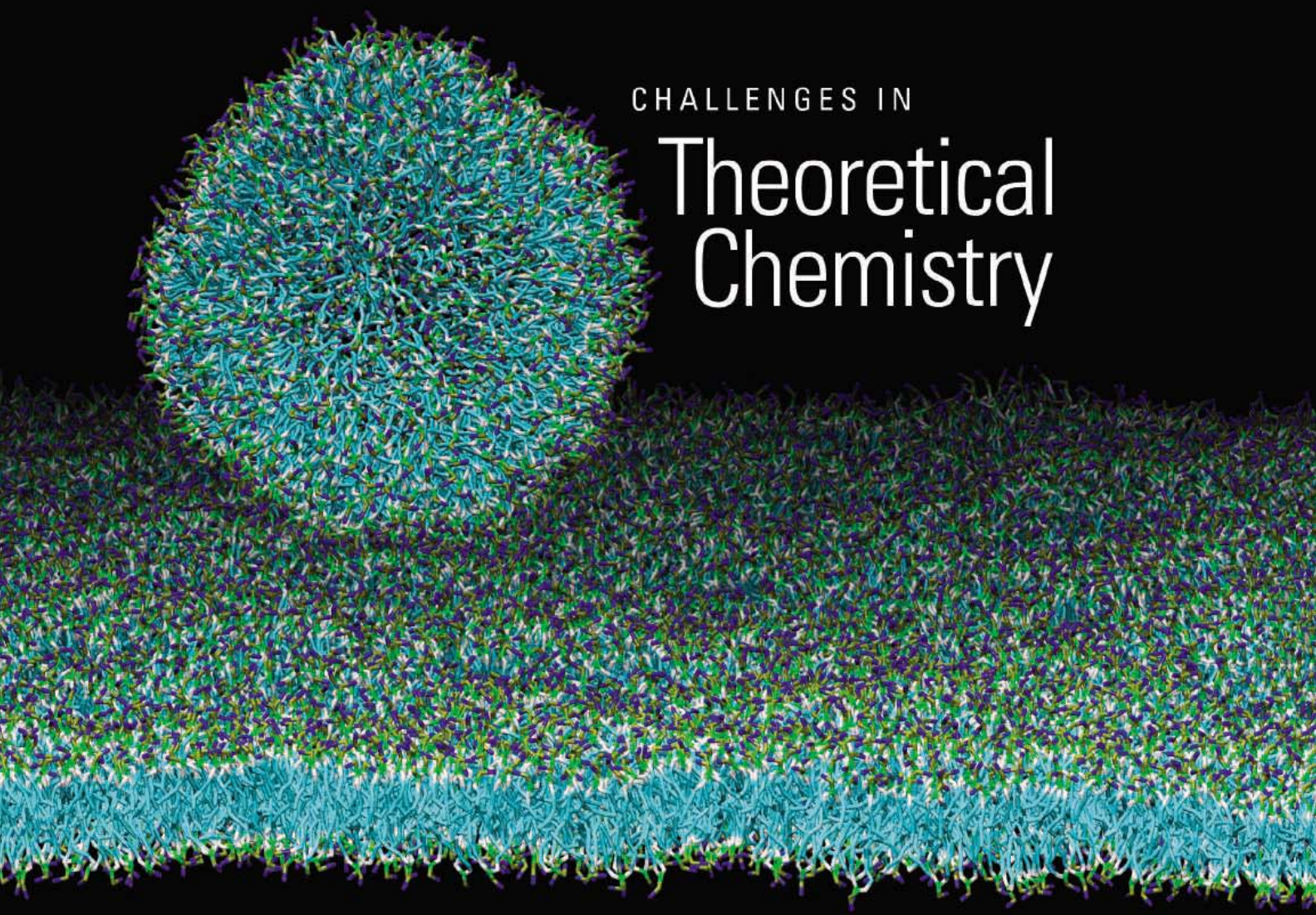
8 August 2008 | $10

Science

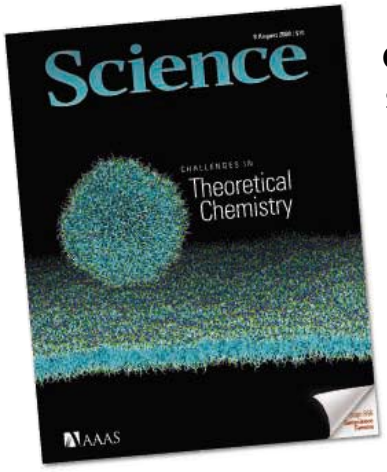


CHALLENGES IN Theoretical Chemistry

page 856
Geoscience Careers

AAAS

Volume 321, Issue 5890



COVER

Simulation of a vesicle interacting with a lipid bilayer (lipid head groups in green and blue, tails in cyan; water is not shown). The computation uses a coarse-grained model with over 1 million interaction sites, equivalent to more than 10 million atoms. See the special section beginning on page 783.

Image: Wataru Shinoda/Research Institute for Computational Sciences, National Institute of Advanced Industrial Science and Technology, Japan; Mike Klein/Center for Molecular Modeling, University of Pennsylvania

DEPARTMENTS

EDITORIAL

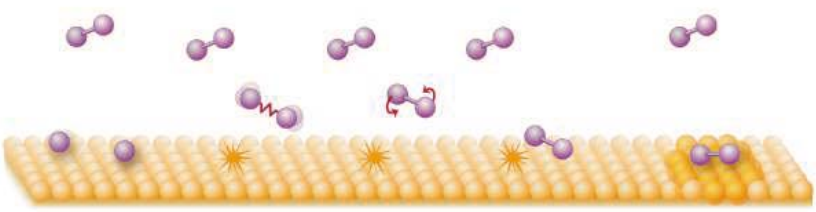
SPECIAL SECTION

Challenges in Theoretical Chemistry

INTRODUCTION

NEWS

PERSPECTIVES



794

760

NEWS OF THE WEEK

NEWS FOCUS

CONTENTS continued >>

SCIENCE EXPRESS

www.sciencexpress.org

ATMOSPHERIC SCIENCE

Atmospheric Warming and the Amplification of Precipitation Extremes
R. P. Allan and B. J. Soden
Satellite data show that in the tropics, heavy rain events have increased in warmer months and decreased in colder months, more than predicted by climate models.
10.1126/science.1160787

BIOCHEMISTRY

Helical Structures of ESCRT-III Are Disassembled by VPS4
S. Lata et al.
A protein responsible for the final separation of daughter cells or budding viruses forms heteromeric complexes on the inside of the membrane to regulate the abscission step.
10.1126/science.1161070

MATERIALS SCIENCE

A Rubberlike Stretchable Active Matrix Using Elastic Conductors
T. Sekitani et al.
A carbon nanotube–polymer film containing organic transistors and coated with silicon rubber can maintain its electrical properties while being stretched up to 70 percent.
10.1126/science.1160309

EVOLUTION

Parasite Treatment Affects Maternal Investment in Sons
T. E. Reed et al.
Mother seabirds that are infected by parasitic nematodes are less able to gather food and feed their fast-growing sons, shifting the sex ratio and affecting population viability.
10.1126/science.1159466

LETTERS

BOOKS *ET AL.*

POLICY FORUM

PERSPECTIVES

BREVIA

CHEMISTRY

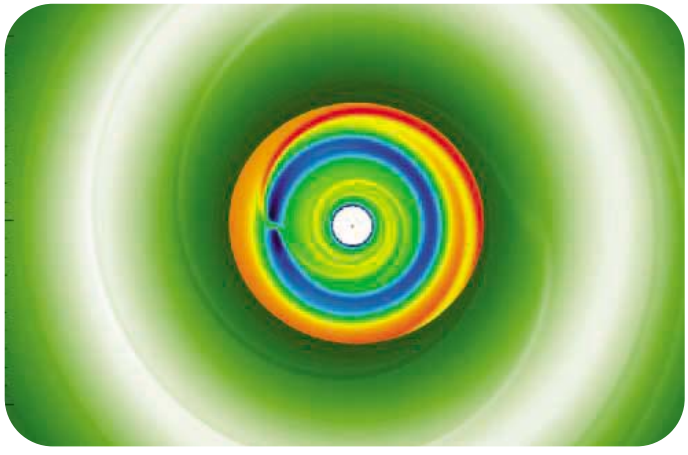
RESEARCH ARTICLES

PSYCHOLOGY

BIOCHEMISTRY

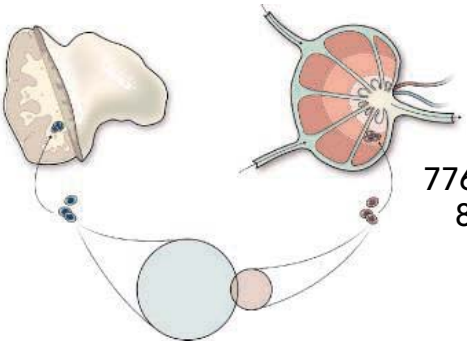
REPORTS

ASTRONOMY



777 & 814

CONTENTS continued >>



776 & 843

REPORTS *CONTINUED...*

SPECIAL FEATURE

Geoscience Careers

www.sciencecareers.org

>> Science *Podcast*

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals Mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices.

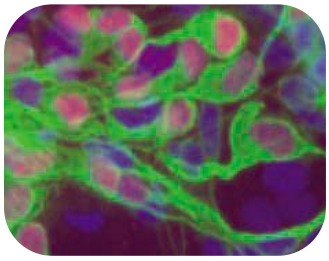
Domestic individual membership and subscription (51 issues): $144 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $770; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery) $85. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #1254 88122. Publications Mail Agreement Number 1069624. SCIENCE is printed on 30 percent post-consumer recycled paper. **Printed in the U.S.A.**

Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $10.00 current issue, $15.00 back issue prepaid includes surface postage; bulk rates on request.

Science is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

Printed on 30% post-consumer recycled paper.

CONTENTS continued >>



Signs of neurons.

SCIENCE**NOW**

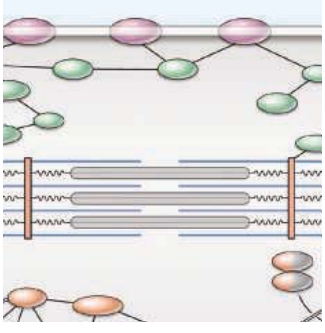
www.sciencenow.org

HIGHLIGHTS FROM OUR DAILY NEWS COVERAGE

Poached Tusks Point to Killing Fields
Genetic analysis enables researchers to trace contraband to region of origin.

Unmasking Dark Energy
Astronomers find its unseen hand at work in the biggest structures in the universe.

Stem Cell Breakthrough in ALS Research
Researchers reprogram skin cells from a patient to become motor neurons.



Sarcomeric proteins.

SCIENCE **SIGNALING**

www.sciencesignaling.org

THE SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

PERSPECTIVE: The Dynamic Z Bands of Striated Muscle Cells
J. M. Sanger and J. W. Sanger
In contrast to its stolid image, the Z band of the mature myofibril is a beehive of activity.

FUNDING SOURCES
Browse a list of grants and funding opportunities for cell signaling research and training; in the Resources section.

Opportunities for geoscientists.

SCIENCE **CAREERS**

www.sciencecareers.org/career_development

FREE CAREER RESOURCES FOR SCIENTISTS

***Science* Careers Podcast: Geoscience Careers**
K. Travis
Hear from experts and geologists about the current job market for geoscientists.

>> See also Geoscience Careers feature p. 856

SCIENCE**PODCAST**

www.sciencemag.org/multimedia/podcast

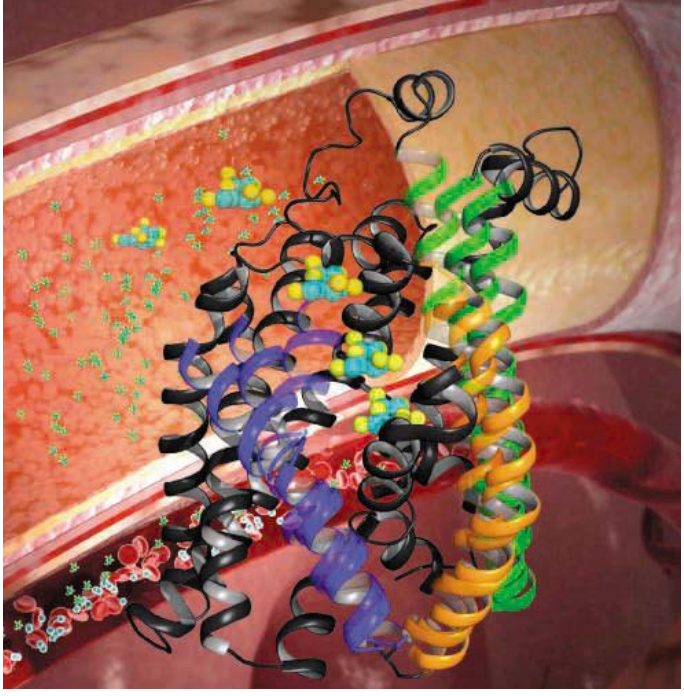
FREE WEEKLY SHOW

Download the 8 August *Science* Podcast to hear about science in Muslim countries, the brain signature of borderline personality disorder, geoscience careers, and more.

CREDITS: (*SCIENCE NOW*) J. T. DIMOS *ET AL.*; (*SCIENCE CAREERS*) NSF/EARTHSCOPE; (*SCIENCE SIGNALING*) CHRIS BICKEL

Separate individual or institutional subscriptions to these products may be required for full-text access.

EDITED BY STELLA HURTLEY



<< Symporter Structure Solved

Solute sodium symporters (SSS) are a large family of proteins that couple the transport of sodium out of the cell with the transport of nutrients into the cell, but the molecular mechanism of this symport has remained unclear. Now **Faham *et al.*** (p. 810, published online 3 July; see the Perspective by **Karpowich and Wang**) have determined the structure of the sodium galactose symporter from ***Vibrio parahaemolyticus*** (vSGLT). The structure obtained is in an inward facing conformation with galactose bound and blocked from exit by a gating residue. Surprisingly the core structure has a similar topology to the core structure of LeuT, a member of the neurotransmitters sodium symporter (NSS) family that has no significant sequence similarity to the SSS family. Based on the LeuT structure an outward facing conformation could be modeled for vSGLT that, together with biophysical data, provides insight into the mechanism of active transport.

A Question of Trust

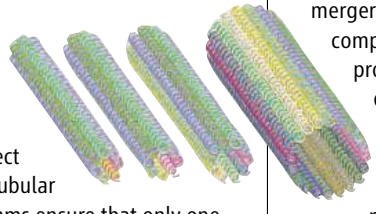
Patients with borderline personality disorder (BPD) have difficulty in maintaining social relationships. **King-Casas *et al.*** (p. 806; see the Perspective by **Meyer-Lindenberg**) studied the behaviors and neural activities of BPD patients participating in an economic exchange game—the so-called Trust game—with healthy partners. Patients were less likely to maintain the level of trust required: as trustees paying back less of the profit that results from a tripling of the investments made by healthy subjects, thus causing the investors to scale back the amount that they would risk. In addition, the patients were less able to repair these breaches of trust—which would require offering disgruntled investors overgenerous payouts to induce a return to a cooperative mode of play. Coupled to these behaviors, neural activity in the anterior insula region of the brain indicated that patients did not seem to process the offer of trust (a high investment) any differently than an expression of distrust (a low investment).

Modeling the Birth of Planetary Systems

A large number of planetary systems have now been discovered around other stars. Many of these contain giant planets in a close orbit, but others contain a different geometry. **Thommes *et al.*** (p. 814; see the Perspective by **Papaloizou**) now present a model of planetary evolution that examines the entire process from the initial formation of planets from a protostellar disk through their subsequent evolution. The model confirms that a wide range of distributions are possible but suggests that the final distribution is particularly sensitive to the viscosity and density of the original protostellar disk. Furthermore, it seems that the production of a planetary geometry like our own solar system is unusual.

Nanoscale Standards

A challenge in making nanoscale materials is being able to control overall dimensions in a manner that also ensures uniformity of the end product. **Yin *et al.*** (p. 824) describe the programming of monodispersed molecular tubes of predefined circumferences from simple single-stranded DNA motifs. Simple programs are written by specifying complementary relationships between ports and nodes, which define the way the strands will connect together into larger tubular structures. The programs ensure that only one diameter tube will form from a set of starting DNA strands, thus guaranteeing the size that forms and allowing for long tubes to form.



Flattening the Spin Landscape

Quantum dots are attractive candidates as the basic building blocks for quantum information processing and quantum computation. However, the material of choice, gallium arsenide, causes problems due to the large and fluctuating magnetic background landscape of the Ga atoms. Each electron is coupled to more than a million background spins, resulting in decoherence that limits the spin lifetime to several nanoseconds. **Reilly *et al.*** (p. 817, published online 10 July) now show that a series of carefully applied voltage pulses can diminish the influence of the background spins thereby allowing the spin lifetime (measured in terms of dephasing time) to be extended to more than a microsecond. This lifetime extension should provide a sufficiently wide window to perform millions of operations on the quantum dots.

Breaking in Three

Though bimolecular reactions are fairly common, it is clear from simple collision probabilities that reactions requiring the simultaneous merger of three different molecules will be comparatively rare. What about the reverse process though? What factors might drive a compound to fly apart into three pieces after an injection of energy? **Savee *et al.*** (p. 826) explore this question at the quantum mechanical level in a combined experimental and theoretical study of *sym*-triazine dissociation. The hexagonal compound, which consists of alternating HC and N sites, breaks apart into three HCN products. A partitioning between two competing pathways could be observed—one in which the bonds all break simultaneously, and one in which the bonds break sequentially—based on the nature of the initially populated electronic excited state.

Mars Rocks

Clay minerals, which contain some water in their structure, have been seen in the oldest areas of Mars. New spectroscopic observations from the Mars Reconnaissance Orbiter, with a resolution as low as 18 meters per pixel, allow mapping of the

Continued on page 743

CREDITS (TOP TO BOTTOM): VINCENT CHAPTAL (ABRAMSON LAB); PENG YIN

Continued from page 741

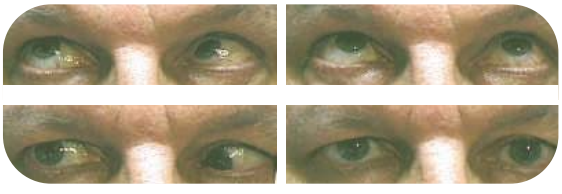
distribution of several clay minerals and reveal a sequence of alteration in one of the oldest outflow valleys of Mars. **Bishop *et al.*** (p. 830) show that the oldest rocks contain abundant smectite rich in iron and magnesium; stratigraphically higher rocks contain abundant kaolinite and montmorillonite, more Al-rich clay minerals, and hydrated silica. These differences may reflect different chemistries of the host rocks or a change in the chemistry and distribution of groundwater over time.

Genetics of the Sexes

Sex determination is a fundamental biological trait in plants, directly coupled to the evolution of mating systems and with tremendous practical significance in fruit and hybrid seed production for many crop species. Changes in the ratio of male and female sexual organs are common in plants, although the underlying genetics are generally not well understood. Andromonoecious plants possess both hermaphroditic and male flowers and have been observed in many different species of plant. **Boualem *et al.*** (p. 836) now show that in melons (*Cucumis melo*) the *andromonoecious (a)* locus encodes the 1-aminocyclopropane-1-carboxylic acid synthase gene (ACS). The CmACS protein is important in the synthesis of ethylene, a plant hormone that influences the development of the plant's sexual organs.

Axonal Pathfinding in Sight

About 1 in 1000 people are afflicted with Duane's retraction syndrome (DRS), a complex congenital eye disorder characterized by a restricted ability to move the eye(s) outward or inward. The condition is thought to arise from faulty innervation of extraocular muscles by cranial motor neurons, which probably occurs early in embryogenesis.



Miyake *et al.* (p. 839, published online 24 July) now provide genetic evidence that strongly supports this hypothesis. Studying families with a variant form of DRS, the authors discover that the mutations responsible for the disorder fall within a gene on chromosome 2 encoding α2-chimaerin, a RacGAP signaling protein previously implicated in axonal pathfinding of corticospinal nerves in mice. The human mutations cause α2-chimaerin to become overactive, and expression of the mutant protein in a chick embryo model did indeed disrupt oculomotor axon development.

Autoimmune Fail-Safe Strategy

The immune system detects and destroys foreign antigens that enter the body; at the same time it must avoid destroying the organism's own antigens, a process that can cause autoimmune disease. To this end, medullary epithelial cells in the thymus express the *Autoimmune Regulator* (*Aire*) gene, which promotes the expression of many of these self antigens. As immune cells mature in the thymus, the ones that recognize these self antigens are deleted. **Gardner *et al.*** (p. 843; see the Perspective by **Kyewsky**) now describe an auxiliary system in the lymph nodes and spleen that ensures that circulating immune cells remain tolerant. To accomplish this, the *Aire* gene triggers the expression of a different array of self antigens in the epithelium of these peripheral lymph organs. There, antigen-specific interactions occur between *Aire*-expressing cells and autoreactive T cells, presumably leading to deletion of any self-reactive T cells that have escaped deletion in the thymus.

Noticing Is Remembering

The dominant model of human visual working memory allows for the simultaneous represention of only three or four objects. With what precision is each visual object stored as a function of the number of items in a scene? **Bays and Husain** (p. 851) tested the ability of human subjects to remember the location and orientation of multiple visual items after a brief disappearance of the stimulus array, and found that visual working memory is a flexibly allocated resource. Making an eye movement toward an object, or directing covert attention to it, caused a greater proportion of memory resources to be allocated to that object, allowing the memory of its presence to be retained with far greater precision than other objects in the scene.

CREDIT: MIYAKE *ET AL.*

Ismail Serageldin is the director of the Library of Alexandria, Alexandria, Egypt.

Science in Muslim Countries

WITH MORE THAN A TRILLION DOLLARS IN CASH AND A POPULATION OF OVER A BILLION PEOPLE, the Muslim world should be poised for a remarkable scientific explosion. Yet despite some very high-profile projects in the Gulf, including the building of massive state-of-the-art facilities for research across all disciplines (and serious efforts elsewhere), the reality is that Muslim countries tend to spend less on scientific research itself, as distinct from buildings and equipment, as compared to other countries at the same income scale. Furthermore, even where funding for science has been available, the results in terms of output—research papers, citations, and patents—are disappointingly low. Why?

Throughout the Muslim world, we are witnessing an increasingly intolerant social milieu that is driven by self-appointed guardians of religious correctness, who inject their narrow interpretation of religion into all public debates. Rejecting rationality or evidentiary approaches, they increasingly force dissenting voices into silence and conformity with what they consider acceptable behavior. Of course, Muslim zealots are not alone in challenging the scientific enterprise; in the United States, battles over evolution and creationism continue to rage.

Yet it was our Muslim forefathers who first held up the torch of rationality, tolerance, and the advancement of knowledge throughout the dark ages of medieval Europe. Centuries before the European scholars Bacon, Descartes, and Galileo considered the scientific method, the great thinker Ibn Al-Haytham (10th century) laid down the rules of the empirical approach, describing how the scientific method should operate through observation, measurement, experiment, and conclusion, the purpose being to "search for truth, not support of opinions." Likewise, Ibn Al-Nafis (13th century) stressed the importance of accepting contrarian views, subject to the test of evidence and rational analysis.

This is the Muslim tradition that must be revived if current efforts are to bear the scientific fruit that a billion Muslims need and that the world has a right to expect of us. Rejecting politicized religiosity and reviving these traditions would promote the values of science in our societies.

There is a central core of universal values that any truly modern society must possess, and these are very much the values that science promotes: rationality, creativity, the search for truth, adherence to codes of behavior, and a certain constructive subversiveness. Science requires much more than money and projects. Science requires freedom: freedom to enquire, to challenge, to think, and to envision the unimagined. We must be able to question convention and arbitrate our disputes by the rules of evidence. It is the content of scientific work that matters, not the persons who produced it, regardless of the color of their skin, the god they choose to worship, the ethnic group they were born into, or their gender. These are the values of science, but even more, they are societal values worth defending, not just to promote the pursuit of science but to have a better and more humane society.

The future can be bright, but it requires a commitment to fight for the values of science and to reject obscurantism, fanaticism, and xenophobia. It requires that members of the scientific and academic communities in Muslim countries be willing to challenge accepted populist views and insist on creating the "space of freedom" necessary for the practice of science and the advancement of knowledge. We must engage with the media and the public and defend the values of science in our societies. These efforts will not be easy, but they constitute a major and necessary step toward liberating minds from the tyranny of intolerance, bigotry, and fear, and opening the doors to free inquiry, tolerance, and imagination.

– Ismail Serageldin

10.1126/science.1162825

CHEMISTRY

Tolerating Chlorine

Supplies of fresh water are steadily dwindling, but salt water remains plentiful, and desalination is increasingly being used for purification. Membrane-based desalination methods require less energy than do distillation-based approaches and are now the dominant technology. However, a complex pretreatment protocol is necessary. Feed waters must be treated with chlorine to eliminate microorganisms that would deposit biofilms onto the membranes, but the chlorine must subsequently be removed to prevent chemical damage to the membranes. After passing through the membranes, the water is then rechlorinated before it is distributed for use. These dechlorination and rechlorination steps increase water purification costs. Park *et al.* have now developed membranes that can tolerate chlorine over a wide pH range. The membranes consist of disulfonated copolymers, which retain the desirable properties of polysulfone—a tough and stable thermoplastic—but are less hydrophobic. The membranes have the potential to be tailored for particular uses and should not require dechlorination of feeds. Further work is required to optimize water transport rates through, and salt retention by, these membranes. — JFU

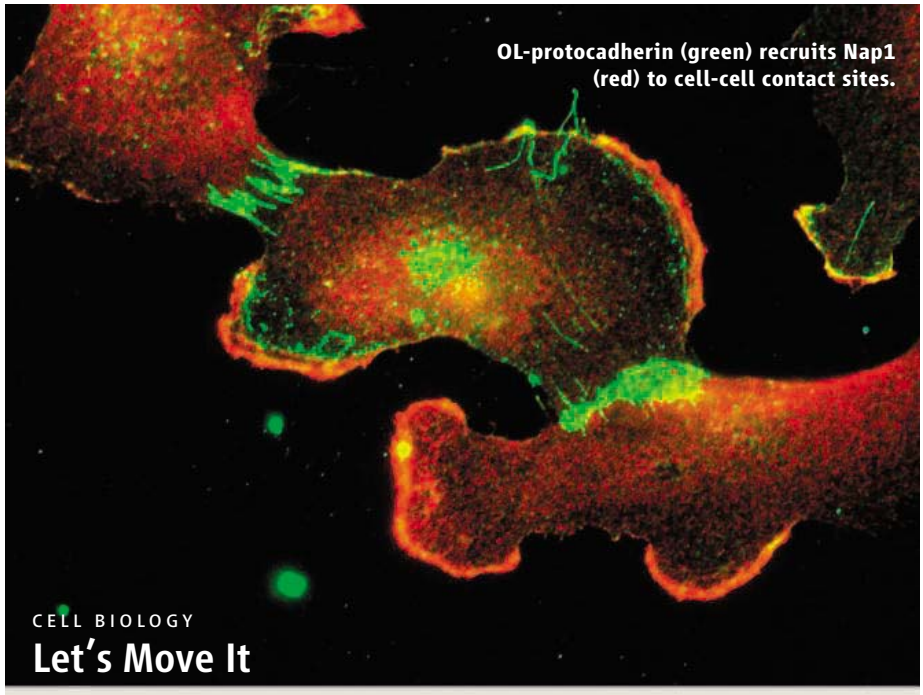
Angew. Chem. Int. Ed. **47**, 6019 (2008).

APPLIED PHYSICS

Unfolding the Power of Solar

Thin-film photovoltaics, such as those based on amorphous silicon or organic films, can be deposited over large areas and so offer the potential to provide a cheap power source that harnesses the free energy from the Sun in which we bask. Combined with the ease of deposition onto flexible substrates, these films also offer the possibility of a lightweight portable power source suited to installation in remote areas. However, the conversion efficiency of such solar cells is relatively low compared to that of their single- and polycrystalline silicon cousins, the workhorses of the present "renewables" technology base for electricity generation. A larger-area film-based cell would thus be required to produce the same amount of power, which could hamper the above-mentioned keenly sought-after applications. Zhou *et al.* now show that the flexibility of thin-film polymer solar cells can get around this problem. They demonstrate a polymer solar cell that can be unfolded like a map. The V-shaped corrugations of the unfolded cell not only enhance practicality but also serve to optimize the collection of light (by multiple reflection) so that the overall efficiency of the cell increases. The future prospects for these thin-film solar cells have just gotten a little bit brighter. — ISO

Appl. Phys. Lett. **93**, 33302 (2008).



OL-protocadherin (green) recruits Nap1 (red) to cell-cell contact sites.

CELL BIOLOGY

Let's Move It

Animal morphogenesis requires that both individual cells and groups of them migrate in a concerted fashion. The mechanisms involved in migration and coordination have been the focus of many studies. Nakao *et al*. reveal how a protein with a role in cell-cell contact actually promotes the joint migration of interacting cells. Generally, cell-cell contact is a signal for cells to stop moving. However, the interaction between OL-protocadherin molecules on neighboring cells stimulates the Nap1-dependent and WAVE-dependent rearrangement of the actin cytoskeleton and encourages the cells to move in tandem. Using astrocytoma cells that expressed OL-protocadherin, or Nap1, or both, the authors defined a pathway through which OL-protocadherin specifically stimulates joint cell migration, while having no effect on an individual cell's capacity to migrate. — SMH

J. Cell Biol. **182**, 395 (2008).

BIOCHEMISTRY

Ready, Set, Go

Synaptotagmin 1 (Syt1) is a Ca^{2+}-binding protein in synaptic vesicles and triggers rapid exocytosis (vesicle fusion with the plasma membrane) in response to Ca^{2+} influx. Fusion mediated by Syt1 and SNAREs (a family of membrane fusion proteins) can be studied in vitro by mixing two populations of vesicles that have been reconstituted with SNAREs: one population with target membrane–associated SNAREs and one with the synaptic vesicle SNARE synaptobrevin 2. Addition of the soluble cytoplasmic domain of Syt1 in the presence of Ca^{2+} triggers fusion, whereas addition of the soluble cytoplasmic domain of synaptobrevin (cd-syb) immediately blocks fusion through competitive inhibition. Chicka *et al.* use this assay to show that both the rate and extent of Ca^{2+}-triggered fusion are increased when the vesicle mixture is pre-incubated with Syt1 for 20 min before addition of Ca^{2+}. Fur-

*Nilah Monnier is a summer intern in *Science*'s editorial department.

CREDITS (TOP TO BOTTOM): NAKAO *ET AL.*, *J. CELL BIOL.* **182**, 395 (2008); ZHOU *ET AL.*, *APPL. PHYS. LETT.* **93**, 33302 (2008)

thermore, fusion is no longer immediately inhibited by cd-syb when vesicles have been pre-incubated with Syt1, demonstrating the presence of a population of docked vesicles in which SNAREs from opposing membranes are already paired. Syt1, therefore, acts in the absence of Ca^{2+} to increase the number of docked fusion-ready vesicles, possibly by stalling partially assembled SNARE complexes. In vivo, this function of Syt1 may contribute to the buildup of docked vesicles, which is essential for the rapid and coordinated release of neurotransmitter. — NM*

Nat. Struct. Mol. Biol. 10.1038/nsmb.1463 (2008).

CHEMISTRY

Stringing Large Rings

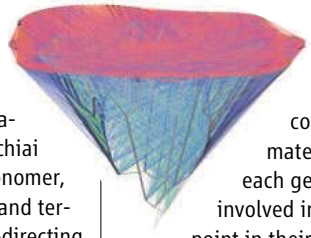
Polymerization reactions that yield products with large cyclic side chains tend to be rare and often exhibit high product polydispersity (a measure of the spread in the length of chains created). However, large cyclic side chains can act as sites for trapping small molecules, which make the polymers useful as absorbents or stationary supports in chromatography, Ochiai *et al*. report that a *bis*-methacrylate monomer, in which a central cyclohexane linkage and terminal urethane groups act as structure-directing agents, can be used to form polymers with 19-atom cyclic side chains; the macrocycles close during the chain-propagation process. Low polydispersity was observed for RAFT synthesis (reversible addition–fragmentation chain-transfer polymerization, a form of living chain polymerization) in dioxane solvent with cumyl dithiobenzoate as a RAFT agent. The controlled nature of the polymerization should allow further modifications through the formation of block copolymers. — PDS

J. Am. Chem. Soc. **130**, 10.1021/ja801491m (2008).

EVOLUTION

Taking the Long View

It can be difficult to establish the phylogeny of microorganisms because they are composed of genes that have moved vertically (via inheritance) or horizontally (via lateral transfer mechanisms such as conjugation) or both. Dagan *et al.* have applied a network analysis approach to estimate the cumulative impact of lateral gene transfer in the genomes of 181 fully sequenced prokaryotes. By examining the presence or absence of all genes and by tracing the evolutionary history of these genes on the basis of genome size, they were able to calculate the rate of lateral gene transfer and have concluded that approximately 80% of the genes in each genome appear to have been involved in lateral transfer at some point in their history. Hence, well-defined phylogenetic trees, which describe genetic relationships accurately on short-term evolutionary time scales, become rather less clearly delineated when looked at over very long time periods. — LMZ



A minimal lateral network.

CREDIT: DAGAN *ET AL.*, *PROC. NATL. ACAD. SCI. U.S.A.* **105**, 10039 (2008)

Proc. Natl. Acad. Sci. U.S.A. **105**, 10039 (2008).

Science Signaling
AAAS

<< Just Looking or Settling In?

Natural killer (NK) cells navigate to transformed or virus-infected cells and bind to them through integrins and NK receptors to form a lytic synapse. Both steps depend on the actin cytoskeleton, leading Butler *et al*. to investigate the role of HS1 (a homolog of the actin-binding protein cortactin) in NK cell–mediated cytolysis. When NK cells were exposed to target cells or to beads coated with ICAM-1 (a ligand of the β2 integrin LFA-1) and the NK receptor ligand ULBP, HS1 localized to the contact site and became phosphorylated on tyrosine. Experiments in which HS1 was knocked down and cells were transfected with HS1 mutants where one or both of two tyrosine residues were substituted with phenylalanine implicated HS1 phosphorylation in NK cell cytolytic activity. Adhesion to ICAM-1 stimulated phosphorylation of HS1 on Tyr^{397}; further, Tyr^{397} was required for chemokine-dependent conversion of LFA-1 into a high-affinity state and for downstream recruitment of actin, the actin regulator WASp, and the guanine nucleotide exchange factor Vav1 to the lytic synapse. Although HS1 Tyr^{397} was not required for recruitment of the adaptor DAP10 to the NKG2D receptor, it was implicated in downstream signaling. In contrast, phosphorylation of HS1 Tyr^{378} was required for chemotaxis. Thus, HS1 appears to be critical to NK cell chemotaxis, formation of the lytic synapse, and cytolysis, and may act as a switch to enable NK cells to convert from a migratory mode to one in which they form a stable contact with a target cell. — EMA

Nat. Immunol. **9**, 887 (2008).

www.sciencemag.org

Science

1200 New York Avenue, NW
Washington, DC 20005
Editorial: 202-326-6550, FAX 202-289-7562
News: 202-326-6581, FAX 202-371-9227
Bateman House, 82-88 Hills Road
Cambridge, UK CB2 1LQ
+44 (0) 1223 326500, FAX +44 (0) 1223 326501

Subscription Services For change of address, missing issues, new orders and renewals, and payment questions: 866-434-AAAS (2227) or 202-326-6417, FAX 202-842-1065. Mailing addresses: AAAS, P.O. Box 96178, Washington, DC 20090-6178 or AAAS Member Services, 1200 New York Avenue, NW, Washington, DC 20005

Institutional Site Licenses please call 202-326-6755 for any questions or information

Reprints: Author Inquiries 800-635-7181
Commercial Inquiries 803-359-4578

Permissions 202-326-7074, FAX 202-682-0816

Member Benefits AAAS/Barnes&Noble.com bookstore www.aaas.org/bn; AAAS Online Store http://www.apisource.com/aaas/ code MKB6; AAAS Travels: Betchart Expeditions 800-252-4910; Apple Store www.apple/eppstore/aaas; Bank of America MasterCard 1-800-833-6262 priority code FAA3YU; Cold Spring Harbor Laboratory Press Publications www.cshlpress.com/affiliates/aaas.htm; GEICO Auto Insurance www.geico.com/landingpage/go51.htm?logo=17624; Hertz 800-654-2200 CDP#343457; Office Depot https://bsd.officedepot.com/portalLogin.do; Seabury & Smith Life Insurance 800-424-9883; Subaru VIP Program 202-326-6417; VIP Moving Services http://www.vipmayflower.com/domestic/index.html; Other Benefits: AAAS Member Services 202-326-6417 or www.aaasmember.org.

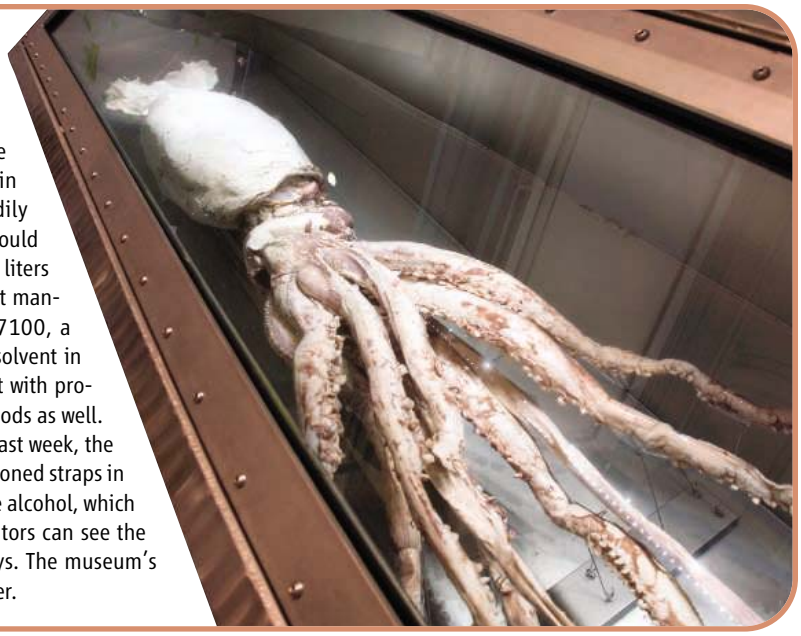
science_editors@aaas.org	(for general editorial queries)
science_letters@aaas.org	(for queries about letters)
science_reviews@aaas.org	(for returning manuscript reviews)
science_bookrevs@aaas.org	(for book review queries)

Published by the American Association for the Advancement of Science (AAAS), *Science* serves its readers as a forum for the presentation and discussion of important issues related to the advancement of science, including the presentation of minority or conflicting points of view, rather than by publishing only material on which a consensus has been reached. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

AAAS was founded in 1848 and incorporated in 1874. Its mission is to advance science, engineering, and innovation throughout the world for the benefit of all people. The goals of the association are to: enhance communication among scientists, engineers, and the public; promote and defend the integrity of science and its use; strengthen support for the science and technology enterprise; provide a voice for science on societal issues; promote the responsible use of science in public policy; strengthen and diversify the science and technology workforce; foster education in science and technology for everyone; increase public engagement with science and technology; and advance international cooperation in science.

Information for Authors
See pages 634 and 635 of the 1 February 2008 issue or access www.sciencemag.org/about/authors

A Squid for the Ages

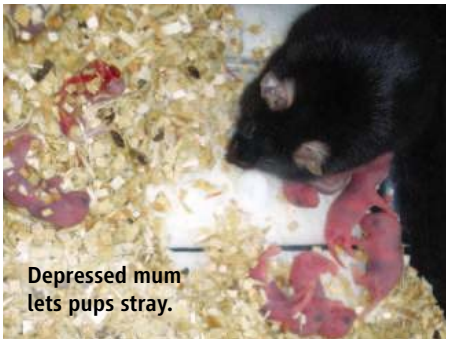


Well-preserved giant squid specimens aren't exactly a dime a dozen. So when Spanish researchers offered to lend intact male and female *Architeuthis* to the Smithsonian Institution in Washington, D.C., for its new Ocean Hall, the staff readily accepted. There was just one problem: A traditional display would have required immersing the 7-meter squid in more than 4500 liters of flammable alcohol. "The fire people went nuts," says project manager Elizabeth Musteen. The solution, literally, was Novec 7100, a hydrofluoroether fluid created by 3M and typically used as a solvent in electronics manufacturing. Because the chemical doesn't react with proteins, company researchers thought it might preserve cephalopods as well.

It did. Installed at the National Museum of Natural History last week, the squids float eerily in their cases, held only by strategically positioned straps in the fluid, which is two and a half times as dense as water. Unlike alcohol, which gives specimens a pale yellow cast over time, it's clear, so visitors can see the brick-red color on the patches of remaining skin, Musteen says. The museum's exhibit on the world's largest ecosystem will open 27 September.

Mouse Mother Blues

Everyone knows mice can get depressed. You can tell if a mouse is depressed if it gives up struggling and just floats shortly after being put in a tank of water, or if it doesn't care whether there's sugar in its drinks. Now scientists say they have a mouse model for another version of the problem: postpartum depression.



Depressed mum lets pups stray.

To avoid anxiety and depression, mice need the right balance between neurosteroids—which leap up during pregnancy—and the so-called GABA neurotransmitter system. Neuroscientists Jamie Maguire and Istvan Mody of the University of California, Los Angeles, observed that the numbers of certain GABA receptors rise and fall in response to changes in steroid hormones over a female mouse's ovarian cycle. In a pregnant mouse, a surge in these neurosteroids causes a decrease in receptors, which normally bounce back after the mouse gives birth.

Mother mice with "deficient" GABA receptors, however, didn't bounce back. Instead, the mothers were sloppy in their nest-building and let their pups scatter and often ate them, the researchers report in the 31 July *Neuron*.

Maguire says the results may indicate that postpartum depression—and premenstrual syndrome (PMS), which often afflicts the same women—have somewhat different biochemical causes from other types of depression. Neuroscientist Nancy Desmond of the U.S. National Institute of Mental Health in Bethesda, Maryland, says the researchers have come up with an "intriguing clue" about what goes awry with depressed new mothers.

The researchers are now testing women with PMS or postpartum depression to see if they have the GABA receptor mutation that predicts vulnerability in the mice.

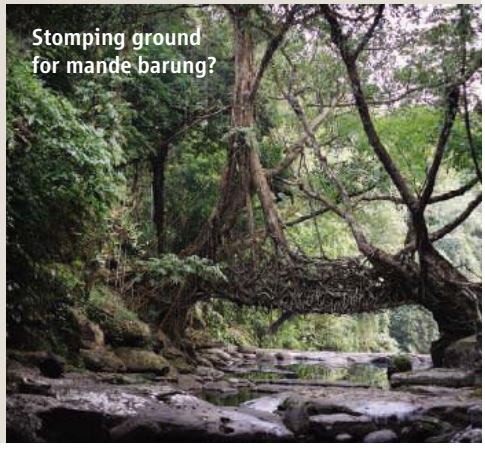
HAIR OF THE YETI

Could two hairs finally solve the mystery of the Yeti? Two small strands sticking from a rock in India are now in the hands of primatologist Ian Redmond and colleagues at Oxford Brookes University in Oxford, U.K.

The hairs are from the hills of Meghalaya, in northeast India. A BBC journalist contacted Redmond, who works with the United Nations on the Great Apes Survival Project, after obtaining samples from Dipu Marak, an Indian naturalist. Marak retrieved them from a crack in a rock after hearing reports of a Yeti-like creature—known locally as "mande barung," or forest man.

Microscopic analysis and comparisons to photographs have ruled out all likely candidates, including the Asiatic black bear, macaque, gorilla, orangutan, Eurasian wild boar, human, and dog. The scientists plan to send each hair to two separate laboratories for DNA analysis to see if they match any known species. "We can't tell from DNA if it's 10 feet [3 m] tall or bipedal, but we can tell if it's related to humans or [other] primates," Redmond said. Another possibility is that the hairs belong to an undiscovered primate species.

"Every scientist in the world would love to believe that there's a big unknown primate out there, but until someone produces concrete evidence, it's impossible to believe," says Craig Stanford, an anthropologist at the University of Southern California in Los Angeles.



Stomping ground for mande barung?

CREDITS (TOP TO BOTTOM): 3M; JAMIE MAGUIRE; TIMOTHY ALLEN/GETTY IMAGES

EDITED BY YUDHIJIT BHATTACHARJEE

CREDITS (TOP TO BOTTOM): WALT DISNEY PICTURES/PIXAR; INSET: UNIVERSITY OF CALIFORNIA, IRVINE; SMITHSONIAN INST.; SOURCE: THE WAKE FOREST & BRIGHAM AND WOMEN'S CENTER; IIHR/UNIVERSITY OF IOWA

Two Cultures

LOOKING AHEAD. James Hicks studies how the human body responds to zero gravity. In 2005, Hollywood's Andrew Stanton, then directing the animated film *WALL-E*, phoned with a question Hicks had never considered: How would 700 years of space travel affect the human race? "They're going to lose muscle mass, get atrophied limbs, and have difficulty getting around—a bloblike phenotype," Hicks, a physiologist at the University of California, Irvine, told the director.

That answer had a big impact on the way Pixar's animators depicted humans in the whimsical sci-fi fantasy that hit theaters this summer. During the making of the film, Hicks met with the crew and delivered a 2-hour talk on the effects of microgravity on the body. The humans in *WALL-E* are bloblike indeed; they rely on personal hovercrafts to move around and are addicted to liquid food and digital entertainment. Some critics have called the movie "antifat," a charge that Hicks calls "ridiculous," noting that the corpulent humans in the film return to redeem a polluted Earth.

IN THE MEDIA

NEVER AGAIN. Floyd Chilton didn't expect to generate much heat with a recent paper on the nutritional content of various farmed fish. But a comparison of farm-raised tilapia with hamburger, bacon, and doughnuts in his discussion has put him at the center of a searing controversy.

Chilton, a biochemist at Wake Forest University School of Medicine in Winston-Salem, North Carolina, studied several types of farmed fish, including the ever-more-popular tilapia. Compared with salmon, tilapia was low in heart-healthy omega-3 fatty acids and high in omega-6 fatty acids. Chilton suspects that this combination can exacerbate inflammation in vulnerable people, such as those with heart disease (a hypothesis that is doubted by other researchers). In the July issue of the *Journal of the American Dietetic Association*, he and his co-authors wrote: "All other nutritional content aside, the inflammatory potential of hamburger and pork bacon is lower than the average serving of farmed tilapia."

Dozens of media outlets ran with the story, some concluding that burgers and tilapia are equally good for the heart. The National Fisheries Institute, a trade group, called the coverage "sensational" and put out a press release with a letter in which 16 researchers and physicians emphasized that swapping hamburger or bacon for tilapia is "absolutely not recommended." Chilton, who feels caught in the crossfire, says he totally agrees. "What I regret is providing a sentence that, although factual, could be such a strong sound bite."

MOVERS

SMART AND LEAN. After 34 years, the Smithsonian Tropical Research Institute (STRI) in Panama is getting a new director: Eldredge "Biff" Bermingham.

The 55-year-old evolutionary biologist will take over what was once a small biological field station and is now considered a premier tropical research organization, with a $25 million annual budget, 40 scientific staff, and about 1000 visiting researchers per year. On staff since 1989, Bermingham has brought molecular studies to bear on tropical systematics and evolution. As deputy director since 2003 and acting director for the past 16 months, he helped departing director Ira Rubinoff expand STRI's work in forest dynamics, including obtaining $8 million in private support to be used in part for a long-term watershed study along the Panama Canal.

Bermingham's expertise in evolutionary biology should help STRI widen efforts to meld ecology and evolutionary biology, says STRI adviser Robert Holt, a professor at the University of Florida, Gainesville. Bermingham hopes that STRI, which gets about two-thirds of its budget from the federal government, can continue to attract the private support it needs to thrive in a tough funding climate. "We're smart, we're lean, and we are willing to take risks," he says.

On Campus >>

EMPIRICAL RESEARCH. When the Iowa River began to flood after a record rainfall in June, hydraulics engineer Larry Weber and his colleagues found themselves both with and without a lab. The rising water forced the researchers to evacuate their offices in the main building of the University of Iowa's Institute of Hydraulic Research in Iowa City. The next day, the surrounding landscape and the river became their backup lab, and studying the flood became a backup project.

The researchers have been collecting water samples and taking measurements of water levels to better understand the flood, which caused about $500 million in damages to the area around the university alone. Later this month, Weber and his colleagues expect to reclaim their offices following cleanup and restoration of the building.

Weber says living with and studying a flood has given him and others at the university a new appreciation for water resources in his native Iowa. "We just kind of take it for granted," he says.

Phoenix's tribulations
758

Eclipse over China
759

BIOTERRORISM

Scientists Seek Answers, Ponder Future After Anthrax Case Suicide

Did he really do it?

That's the main question on the minds of many scientists this week after an Army researcher apparently close to being indicted for the worst bioterror attack in U.S. history took his own life. As researchers tried to make sense of scraps of information filtering out in the media, many were hoping prosecutors would soon reveal their entire case against Bruce Ivins of the U.S. Army Medical Research Institute of Infectious Diseases (USAMRIID) in Frederick, Maryland, so the country can scrutinize the evidence that led the Federal Bureau of Investigation (FBI) to believe he mailed the anthrax-laden letters in fall 2001. That evidence likely includes sophisticated and possibly debatable scientific analyses.

For those who knew Ivins, the question is personal. "He was a nice guy, a sweet guy," says fellow anthrax veteran Martin Hugh-Jones of Louisiana State University in Baton Rouge. "He wasn't on my shortlist of possible suspects." In the wider field, the case is prompting other questions: If one of their own committed the crimes, will the biodefense budget, which ballooned after 2001, shrivel? Will public confidence in the field decline, and will rules for handling possible bioterror agents become draconian?

Ivins, 62, died at a Frederick hospital on 29 July after taking an overdose of painkillers. An author on 54 PubMed-listed papers, he had spent most of his career developing drugs and vaccines against anthrax, studies for which mice, rabbits, and monkeys were frequently exposed to the deadly disease. As *Science* went to press, the FBI had not named Ivins as a suspect, saying more news would be forthcoming after survivors and victims' families had been notified. Ivins's lawyer, Paul Kemp of Rockville, Maryland, has declared his client's innocence, alleging in a statement that mounting FBI pressure had "led to his untimely death."

The case against Ivins is likely to rest on "a combination of investigations—including good old gumshoe work, science, and perhaps other sources of information and evidence," says Randall Murch, a bioforensics expert who worked for the FBI and is now a Virginia Polytechnic Institute and State University administrator based at the university's office in Alexandria.

Case shut? Researchers are clamoring for the FBI to release evidence implicating Bruce Ivins in the anthrax attacks.

In a statement, the FBI credited "new and sophisticated scientific tools" for the "substantial progress" made recently. That, and a report in the *Los Angeles Times* that the bureau recruited Ibis Biosciences, a California company specializing in high-throughput genetic analyses, leads microbial genomicists such as J. Craig Venter of the Venter Institute (JCVI) in Rockville, Maryland, to speculate that an approach called metagenomics—which looks for the genomic makeup of an entire population of cells instead of a single one—may have been used to try to link the *Bacillus anthracis* spores mailed to two U.S. Senators and media outlets to those used in Ivins's lab.

News reports early this week also said genomic analyses suggested that the anthrax powder was a mix of two strains, one obtained at Dugway Proving Ground, a testing facility in Utah, and the other from USAMRIID. Opinions differ sharply among experts about whether a so-called lyophilizer, which Ivins was reported to have used, would suffice to produce the extremely fine, floating powder found in the Senate envelopes.

Whatever the scientific evidence, it would face stiff challenges in court, experts say; in contrast to human DNA traces, whose utility has become well-accepted, microbial forensic evidence is largely untested. Now that Ivins won't face trial, it's even more important that scientists be able to pore over all the evidence, says anthrax researcher R. John Collier of Harvard Medical School in Boston. "I would love to see what they have," Collier says. Still, scientific scrutiny can't replace a court of law, some say. "What's the forum? Are we going to discuss genetic frequencies in a dark hall in a Marriott somewhere?" another anthrax scientist asks.

The questions are critical because the FBI was wrong before. Just 6 weeks ago, the government agreed to pay $5.8 million to Steven Hatfill, a former colleague of Ivins's at USAMRIID whose life was turned upside down in 2002 after then–Attorney General John Ashcroft called him a "person of interest" in the anthrax attacks. Virologist Thomas Geisbert, associate director of Boston University's (BU's) emerging infectious diseases lab, says he can't help wondering whether Ivins's death could be the result of "another Hatfill situation"—except that Ivins was unable to handle the intense pressure. There were signs of mental instability; Ivins had recently been hospitalized for erratic behavior, and on 24 July, a Maryland court issued a restraining order against Ivins at the request of a therapist who said he had a history of making homicidal threats.

The FBI, which had little microbial forensic experience back in 2001, relied on a network of labs—including Ivins's at USAMRIID—to aid its investigation. (The Institute for Genomic Research in Rockville, Maryland, not only sequenced many anthrax strains but worked on the case before it was integrated into JCVI, says Venter.) The bureau has ordered researchers not to discuss or publish that work. "As a scientist, I hope I'll be able to do that now," says Geisbert, who in his previous job at USAMRIID produced electron micrographs of the spores used in the letters sent to the Senate.

Many believe that the case is bound to have wider ramifications for the biodefense field. Before 2001, such research was largely ▶

CREDIT: COURTESY OF *FREDERICK NEWS-POST*

Regulating evolution 760

Biology's big screen 764

confined to USAMRIID and the Centers for Disease Control and Prevention in Atlanta, Georgia. The anthrax letters, which plunged a nation reeling from 9/11 into further anxiety, helped spur a massive increase in the biodefense budget—now some $5.4 billion a year—and a construction boom in biosafety labs. "The entire rationale for that expansion was fraudulent," says Richard Ebright, a prominent biodefense critic at Rutgers University in Piscataway, New Jersey, because it assumed a threat from outside the country. The boom has made the country less safe, Ebright maintains: "The spigot needs to be closed."

Others say the threat remains real. "It would be unfortunate if people take away the message that the only individuals we should be concerned about are deranged biodefense scientists," says biosecurity expert Gerald Epstein of the Center for Strategic and International Studies in Washington, D.C. But he acknowledges that the debate about how much biodefense is enough will likely reignite.

There may be other consequences, says Paul Keim, an expert in microbial fingerprinting at Northern Arizona University in Flagstaff who has also been recruited by the FBI. After the anthrax attacks, Congress passed legislation to limit access to bioterror agents to licensed researchers and imposed strict rules on where and how they can be used. Although researchers have decried them as overly cumbersome, the anthrax case may renew pressure to stiffen them further, Keim says. Additional measures could include cameras in virtually every lab, speculates Alan Pearson of the Center for Arms Control and Non-Proliferation in Washington, D.C. "The solution may well be if you work with pathogens like smallpox and anthrax, be prepared to be watched," he says.

The involvement of a U.S. scientist would also give new ammunition to local groups that have tried to stop construction of new biosafety labs. At BU, now a major academic biodefense hub, "we have had a lot of opposition—and this is not going to help," Geisbert says.

–MARTIN ENSERINK AND YUDHIJIT BHATTACHARJEE

SCIENCE AND THE PUBLIC

Scientists Targeted in California Firebombings

Early Saturday morning, a Molotov-cocktail-like device set fire to the home of a developmental neurobiologist at the University of California, Santa Cruz (UCSC). His family fled down a fire escape from a second-story window. Around the same time, a similar device destroyed the car of another UCSC researcher. As *Science* went to press, no one had claimed responsibility for the attacks, but the university and police suspect they are the work of animal-rights extremists.

Crime scene. Police suspect animal-rights extremists are behind the destruction of a UCSC researcher's car last weekend.

CREDIT: SHMUEL THALER/*SANTA CRUZ SENTINEL*

In recent years, universities and law enforcement officials in the United States have had to grapple with increasingly personal threats, harassment, and attacks on animal researchers and their families (*Science*, 21 December 2007, p. 1856). California has been an epicenter: In the past few years, several biomedical researchers at UC Los Angeles have been targeted, and more recently, scientists at other UC campuses have endured harassment and had their homes vandalized. Twenty-four UC Berkeley researchers and seven staff members have been harassed in the past few months, according to a university spokesperson. In February, six masked intruders tried to force their way into the home of a UCSC researcher during a birthday party for her young daughter.

Concerns were sparked again last week in Santa Cruz by pamphlets discovered in a downtown coffee shop and turned in to police. Titled "Murderers and Torturers Alive and Well in Santa Cruz," they contained the photographs, home addresses, and phone numbers of 13 UCSC faculty members, along with "threat-laden language" condemning animal research, says Captain Steve Clark of the Santa Cruz police.

David Feldheim, the neurobiologist whose house sustained substantial damage in the fire attack, was one of those listed. Feldheim uses mice in studies of how the brain's visual system develops. The researcher whose car was destroyed was not on the list, Clark says. UCSC spokesperson Jim Burns declined to name that researcher or say whether he or she uses animals for research. A third researcher, who was named in the pamphlet, lives "almost next door," Clark says, raising the possibility that the culprits missed their intended target.

On Saturday, UCSC Chancellor George Blumenthal condemned the attacks as "criminal acts of anti-science violence." Several hundred people gathered on campus Monday for a rally in support of the researchers who'd been attacked, according to the *Santa Cruz Sentinel*, and officials announced a $30,000 reward for information leading to the arrest and prosecution of those responsible.

The Santa Cruz police have handed over the investigation to the Federal Bureau of Investigation, which will investigate the incidents as acts of domestic terrorism. "We have some good leads and some helpful witnesses," Clark says, but so far no suspects.

–GREG MILLER

AIR QUALITY

Court Ruling Scrambles Clean Air Plans, Leaving a Vacuum

The U.S. Environmental Protection Agency (EPA), state regulators, and the electric power industry are struggling to come to grips with the impact of a surprise court decision last month that dismantled a major air-pollution regulation. According to EPA estimates, the rules would have prevented more than 13,000 premature deaths by 2010, cut haze, and reduced acid rain, but a federal appeals court, citing fundamental flaws, ordered EPA to scrap the plan. "We're back to square one," says John Walke of the Natural Resources Defense Council in Washington, D.C.

EPA has until 25 August to appeal the decision. Officials there warn that without the new rule, air pollution could worsen, and power companies that proactively implemented pollution controls will in effect be penalized. In a Senate hearing last week, industry representatives and regulators called for Congress to come up with new legislation quickly. But it's unlikely to happen soon, leaving state regulators scrambling to figure out how they will meet air-quality standards.

The regulation, called the Clean Air Interstate Rule (CAIR), was designed to fix a problem faced by many East Coast states: So much pollution blows in from other states that they can't meet EPA standards for air quality. CAIR would have capped the amount of pollution in the entire region and issued each state "allowances" representing units of pollution. Power companies had already started buying and selling these allowances, which provided a financial incentive to clean up their power plants. This "cap-and-trade" scheme was widely seen as a rapid and cost-effective way to clear the air in both upwind and downwind states, similar to the successful scheme enacted to control acid rain in 1990.

CAIR was also based on a trading program for nitrogen oxides (NO_x), which lead to smog. Since 2000, this program, which covers 20 eastern states and operates during the 5-month ozone season, has cut summer NO_x emissions by 60%. It was to end in September as CAIR phased in, although EPA is now mulling whether to extend it.

CAIR was designed to further cut both sulfur dioxide (SO_2) and NO_x emissions over the entire year in 28 eastern states. By 2010, it was projected to lower SO_2 emissions by 45% from 2003 levels and NO_x by 53%. With even tighter caps implemented in 2015, the system was predicted to save up to $100 billion in health and other costs, as well as end chronic acidity in Adirondack lakes by 2030. EPA considers it "the most significant action to protect public health and the environment" in nearly 20 years.

Not everyone was happy, however. Environmentalists and some states complained that the rule was too lax; some sued. Several power companies and states sued EPA for other reasons, including how it distributed the allowances. On 11 July, the U.S. Court of Appeals for the District of Columbia Circuit decided that CAIR had "more than several fatal flaws." Among them, it ruled that CAIR was not stringent enough and that 2015 was too late for tightening the caps.

The loss of CAIR will likely slow efforts to improve visibility in national parks and set back international negotiations over long-range transport of air pollution, predicts Brian McLean, who directs EPA's Office of Atmospheric Programs. And regional pollution could well increase because without cap-and-trade incentives, power companies might not

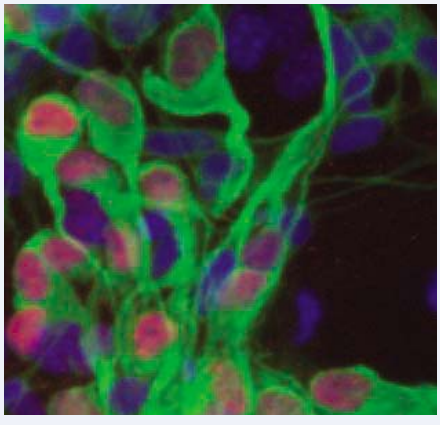
STEM CELLS

Ethics Questions Add to Concerns About NIH Lines

Some federally funded scientists are having second thoughts about working with the 21 human embryonic stem (ES) cell lines available to them under President George W. Bush's policy, following a report indicating that the cells are getting increasingly stale—not only scientifically but ethically as well.

A recent article by Rick Weiss of the Center for American Progress in Washington, D.C., has drawn attention to a paper by bioethicist Robert Streiffer in the May issue of *The Hastings Center Report*. Streiffer, of the University of Wisconsin, Madison, says consent forms signed by embryo donors for the approved lines are inadequate by today's standards. "We know how to do things better now," says Streiffer, who believes this is yet another reason why the Administration's policy, which limits federal funding to work with cell lines derived before August 2001, is untenable.

Streiffer says most of the consent forms fall short of standards for informed consent in embryo research that were in place as early as 1994. And only one, from the University of California, San Francisco, comes close to meeting 2005 guidelines from the National Academy of Sciences. He singles out forms from two companies—BresaGen, now owned by Novocell, and Cellartis—as particularly inadequate. BresaGen's had only one sentence saying that defective embryos created from in vitro fertilization might be used in research. Cellartis told donors that cells would be destroyed after a research project. Other forms failed to mention that embryos would be destroyed and that cells derived from them could end up in experiments around the world.



ES cell alternative? iPS cells have been used to create motor neurons (*above*).

CREDIT: J. T. DIMOS *ET AL.*, *SCIENCE*

"I was shocked," says Lorraine Iacovitti, a neurologist at the Thomas Jefferson University Medical College in Philadelphia, Pennsylvania, who has used one of the BresaGen cell lines. Most researchers "just assumed that the consent had been taken care of." Now two universities, Stanford and Johns Hopkins, are considering prohibiting work with the companies' five cell lines, which are not widely used.

Story Landis, chief of the stem cell task force at the National Institutes of Health (NIH), says no changes are planned in response to the report. Allan Robins of Novocell in Athens, Georgia, says NIH officials told BresaGen in 2001 that "they felt what we had done was reasonable." He says that ideally, the company would ask the donors for

Uncapped. Smog and acid rain may increase if power plants turn off scrubbers.

add or turn on pollution-control equipment, which is expensive to install and operate. State regulators hope to encourage companies to keep the equipment running, but "there is no guarantee that will happen," Christopher Korleski, who heads Ohio's Environmental Protection Agency, told a Senate Environment and Public Works subcommittee.

The court decision penalizes companies that have already spent billions on new equipment to prepare for CAIR, McLean told the Senate subcommittee. PPL Corp. in Allentown, Pennsylvania, for example, has invested nearly $1.5 billion in scrubbers, driven largely by the expected market value of pollution allowances. Those values collapsed after the ruling, and PPL lost roughly $70 million on the SO_2 allowances it had banked. Companies also fear that states will force them to use their new equipment while dirtier competitors won't bear those costs. "That will have a chilling, long-term effect" on investment in pollution-control technology, predicts Eric Svenson of the Public Service Enterprise Group, an energy company in Newark, New Jersey.

Everyone agrees that something needs to be done. Twelve states have asked EPA to repromulgate a CAIR-like rule acceptable to the courts. But PPL Executive Vice President William Spence predicts that any regulatory solution "will continue to be plagued by litigation." EPA's McLean says the agency is evaluating its options. The Senate subcommittee chair, Tom Carper (D–DE), plans to have more hearings this fall, but Senator George Voinovich (R–OH) said he doubts the Senate will deal with the issue until late spring.

–ERIK STOKSTAD

retroactive consent, but it is impossible to trace them. A representative from Cellartis told *Science* the company is preparing a correction to Streiffer's article.

Many scientists say they would prefer to work with new human ES cell lines rather than any of the aging lines on the presidentially approved list. In addition to the ethical concerns, the cells are problematic for scientific reasons—for one, they were grown on mouse "feeder" cells, which makes them unsuitable for use in human therapies. But some scientists have been constrained from switching to new lines because they would lose federal funding.

Two pending developments may change that. Both senators Barack Obama and John McCain have said that they support congressional efforts to expand the number of cell lines available to federally funded researchers. If the new president doesn't act fast enough, Congress likely will; both houses have twice passed such legislation only to be thwarted by Bush vetoes.

In addition, remarkable progress with a new type of cell—induced pluripotent stem (iPS) cells—promises an alternative to cell lines derived from embryos. When Japanese researcher Shinya Yamanaka announced 2 years ago that he had cultivated colonies of ES-like pluripotent cells by inserting just four genes into mouse skin cells, many thought it would be years before the same could be done with human cells. But last year, two groups pulled it off (*Science*, 1 February, p. 560).

In the past 4 months, scientists have used iPS-derived cells to treat blood and neurological disorders in rats and mice, for instance. Two groups at Harvard University have developed colonies of iPS cells from patients with a variety of genetic diseases. Yet other work has shown that small molecules can be substituted for some potentially cancer-causing genes used to derive the original iPS cells.

Major hurdles remain. It's not clear whether iPS cells will behave exactly like ES cells. And they can't be used therapeutically because the viral vectors scientists use to introduce genes could be hazardous. But given the speed of developments, at least one stem cell researcher, Rudolf Jaenisch of the Massachusetts Institute of Technology in Cambridge, believes "we will solve this much earlier than we think."

–GRETCHEN VOGEL AND CONSTANCE HOLDEN

CREDIT: PHOTOS.COM

SCIENCESCOPE

Obama Banks on NASA

Just days after NASA celebrated its 50th birthday, Democratic presidential candidate Barack Obama told a cheering crowd at Brevard Community College near the agency's Kennedy Space Center that he supported the shuttle replacement program and that "we've got to make sure that the money going into NASA for basic research and development continues to go there." Republican candidate John McCain's staff has questioned Obama's support for the agency by noting his proposal last year to pay for $18 billion in new education programs in part by deferring funding for the shuttle replacement (*Science*, 1 February, p. 565). Meanwhile, space representatives from nine countries meeting at NASA's Ames Research Center in Mountain View, California, last week agreed to plan a series of fixed and roving science stations on the moon starting in 2013.

–ANDREW LAWLER

Psychiatrist Dropped From Grant

Under congressional pressure, Stanford University is temporarily pulling a faculty member off a National Institutes of Health (NIH) grant involving a company in which he owns millions of dollars in stock. The company, Corcept Therapeutics, is testing the drug mifepristone as a treatment for depression, and Alan Schatzberg is principal investigator on a multipart NIH grant that includes a mifepristone depression study. Although Stanford says Schatzberg had reported his stock and was not involved with the trial, university officials last week told NIH that they "can see how" the situation "may create an appearance of conflict of interest." U.S. Senator Chuck Grassley (R–IA) has been investigating the broader issue in U.S. universities (*Science*, 27 June, p. 1708). **–JOCELYN KAISER**

Survey Finds More Apes

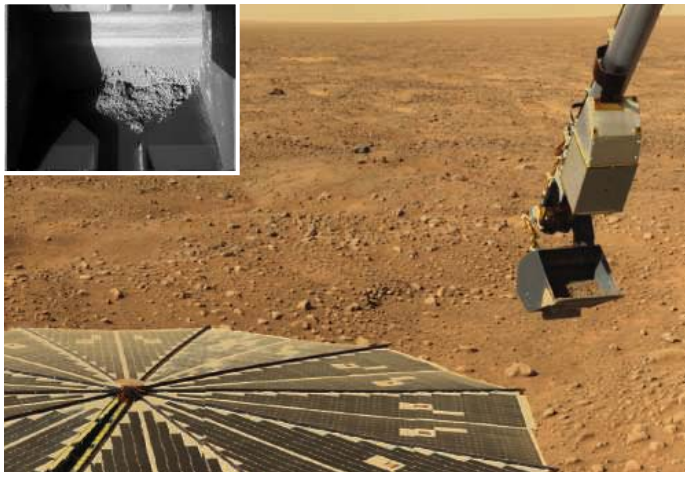
Scientists hope that news of a larger-than-expected population of gorillas in the Republic of the Congo will reinvigorate efforts to protect the critically endangered species (*Science*, 14 September 2007, p. 1484). The Wildlife Conservation Society reported this week that there are 125,000 western lowland gorillas in an area of northern Congo, dwarfing the previous guess of 50,000. The survey, which covers about 10% of the species' range, encompasses areas that the Congolese government has slated for protection. But in recent years, those efforts have stalled. "We hope these results will help catalyze that process," says the society's Emma Stokes. The good news is tempered, however, by a recent Ebola outbreak nearby.

–GRETCHEN VOGEL

PLANETARY SCIENCE

Phoenix's Water May Be Gumming Up the Works

Many media outlets hailed the Phoenix lander last week for confirming the presence of water on Mars. Actually, the two Viking spacecraft won that honor 3 decades ago when they confirmed that the northern polar cap contains water ice. Then rumors began flying that Phoenix has evidence of "potential habitability" on polar Mars, which no one was yet ready to discuss. But the mission's most dramatic achievement so far (see sidebar for others) has been touching martian water ice—which may also be creating the mission's biggest challenge.



Out, damned dirt. The Phoenix robotic arm (*above right*) can scrape up enough dirty ice for a sample, but it won't fall out of the scoop (*inset*).

The often icy soil that Phoenix was sent to analyze "has very interesting physical properties," as Phoenix team member William Boynton put it last week—so interesting that team members spent the middle third of the mission trying in vain to get an ice-rich sample into one of the lander's two prime analytical instruments. Now Phoenix has moved on to less challenging, less icy samples while team members try to sort out the mysteries of alien dirt.

This isn't Phoenix's first problem with martian soil. When the lander's robotic arm dumped a scoop of non-icy near-surface soil onto a screen leading to the Thermal and Evolved-Gas Analyzer (TEGA), the soil just sat there. After scientists vibrated the screen a half-dozen times over several days, the soil just as mysteriously relented and suddenly fell through the screen and filled the TEGA sample cell.

"Right now, I can say none of us knows" what's going on with Phoenix soil, says team member Douglas Ming of NASA's Johnson Space Center in Houston, Texas. Ideas abound, though. The clumpiness—seen on lander and rover missions since the Viking days—might reflect either a buildup of electrostatic charge on the finest particles, a mechanical interlocking at particle edges, or the dampening effect of salts, says Ming. Whatever the cause, lander operators have sidestepped the soil clumpiness by having the arm slowly sprinkle soil from the scoop. In the meantime, however, the extended vibration of the screen apparently caused an electrical short circuit, which forced team scientists to consider that their next TEGA analysis might be their last.

Under pressure to get results, team members went for the gold: the rock-hard dirty ice at the bottom of a 5-centimeter-deep trench dug through the soil. Only ice could yield the isotopic composition of martian water, and it might have preserved much-sought organic matter. Day after day, the Phoenix team practiced how best to scrape, rasp, and scoop up ice chips and deliver the sample to TEGA. Daily visual checks at each step dragged the process out to 30 days, one-third of the planned mission. On the first attempt to deliver a sample, the filled scoop was tilted over TEGA and vibrated, "and nothing came out," says robotic arm co-investigator Raymond Arvidson of Washington University in St. Louis, Missouri. "We repeated the experiment, but with more vibration, and it still didn't come out," even with the scoop turned upside down. "Of all the things that could go wrong, that was the least likely," says Boynton.

Why it went wrong remains a mystery. Planetary scientist David Paige of the University of California, Los Angeles, shares team members' suspicions that at least part of the problem is that Phoenix forcibly removed the ice from the coldest spot around using a relatively warm scoop in a relatively warm atmosphere. That could lead to melting and refreezing, says Ming, much as ice cubes fresh from the freezer can fuse into a single, pesky clump in your glass. Ming says successful tests of the scoop before launch didn't include such changes in temperature. "Maybe we needed to do more testing," he says, but neither time nor funding would have allowed that.

"It's unfortunate we spent 30 [days] working on delivering ice," says Arvidson. That would have left less than 30 days in the planned mission with six TEGA sample cells remaining, each requiring 7 days to analyze. "I'm concerned but not panicked," he adds. NASA has now extended the mission by 30 days, and because plenty of science remains to be done on dry soil, Arvidson says, "we have to get on with business" while they work the icy-soil problem. **–RICHARD A. KERR**

Successes, Past and Future

Now just past the halfway point of its 120-day mission, Phoenix has run into problems handling the martian soil it was sent to analyze (see main text). But it has had its accomplishments, including:

- A successful landing—All the testing of hardware and software inherited from the ill-fated Mars Polar Lander (*Science*, 9 May, p. 738) paid off in a perfect arrival for Phoenix, in part because the landing site turned out to be as safe as scientists had predicted.
- Ice in easy reach—Scientists had deduced from orbital observations and theory that ice would be found anywhere they looked beneath 2 to 6 centimeters of loose soil. Phoenix found it 5 centimeters down on its first try.
- Instrumentation that works—Both major instrument packages yielded results on their first tries. This martian soil, at least, is alkaline, not acidic as expected, and contains the products of interaction with water, although when and where that interaction occurred remains unknown.

Nevertheless, the high-profile mission goals remain elusive. Signs that life may have been possible when the ice melted some time in the geologic past would most likely come from the wet-chemistry analyzer. Rumors that such signs have in fact been detected were rife at press time. The "bake test" analyzer can detect organic matter—the remains of life or merely meteoritic debris—but results from the first sample are requiring weeks of analysis. **–R.A.K.**

ECLIPSE

Researchers Flock to View Fleeting Display of Solar Corona

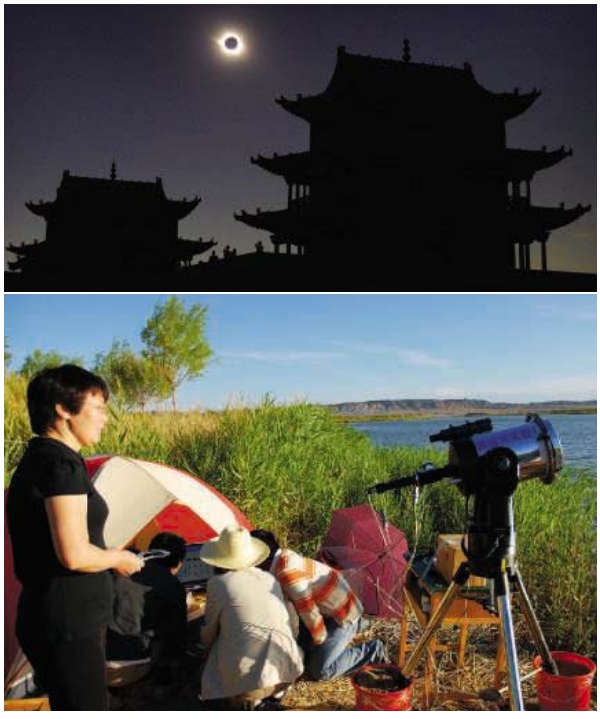
JINTA, CHINA—With anticipation growing by the second, teams of Chinese scientists in matching T-shirts bearing the logos of their institutions fussed over telescopes, cameras, spectrometers, and other instruments set along the rim of a lake. At 7:14 p.m. Beijing time on 1 August, about an hour after the moon began to slide across the sun's face, the blue sky above this town in western China darkened like a sunset in fast motion. Then totality: The moon blotted out the solar disk, leaving the wispy corona, along with Mercury and Venus, visible to the naked eye.

For the 110 seconds of total solar eclipse over Jinta, some of the dozens of scientists gathered here snapped photos while others silently took in the ethereal scene. Then the sky brightened. "This is my first time. It was just fantastic," said Yan Yihua, a solar physicist with National Astronomical Observatories, Chinese Academy of Sciences (NAOC), in Beijing who was recording broadband radio emissions from the corona.

A total solar eclipse is not just an awesome spectacle—it's also a rare opportunity to observe the corona, a swirling halo of plasma that's a millionth as bright as the solar disk. Such studies have taken on added significance in the wake of recent finds from Hinode, a spacecraft that has brought researchers to the threshold of solving a pair of long-standing enigmas: What impels the solar wind, and how the corona is heated to several million kelvin, enormously hotter than the sun's surface (*Science*, 7 December 2007, p. 1571). "Hinode has shown that the solar atmosphere is much more dynamic than we thought," says Kazunari Shibata, a solar physicist at Kyoto University in Japan. But just how the corona and its magnetic field are energized is still largely a mystery—one that experiments during a total eclipse could help shed light on.

China's record of observing the solar eclipse dates to roughly 2000 B.C.E. "In ancient China, people venerated the sun. They thought the solar eclipse is unlucky," says NAOC's Han Yanben. Eclipse observations were vital for checking the calendar, and rulers planned around the unsettling events. For certain officials, failure to observe an eclipse was a grave mistake. During the Xia Dynasty some 4 millennia ago, annals show, court astronomers Xi and He were drunk and missed an eclipse. By law in those times, says Han, they were executed.

The scientists in Jinta were not under that kind of pressure, but the stakes were nonetheless high to get their measurements right. A group from Yunnan Astronomical Observatory in Kunming had set a 20-centimeter telescope hard up against the lake to minimize thermal noise. Minutes after the eclipse, team leader Liu Zhong was hunched over a laptop in a tent next to the telescope. "We can see fine structure here," Liu said, pointing to grainy features just above the sun's limb. "I don't know what it is yet—but it's so good!" he exclaimed. Zhang Mei, an NAOC solar physicist, was impressed. "This is why we gave him the best location," she said.



Show time. Zhang Mei looks on as Liu Zhong (in straw hat) examines early data. Eclipse over Jiayuguan Fort in Gansu Province (*top*).

Uphill, three teams were poring over spectral emissions. In a dark-green tent, Bao Xing-Ming and colleagues from NAOC were using a charge-coupled device (CCD) camera and spectrometer to zero in on the near-infrared. Features of the corona's magnetic field can be deduced from these spectra and their polarization, says Zhang. "We know the corona's magnetic field is important in space weather, but we can't really measure it," she says—except during an eclipse or using a coronagraph that mimics an eclipse. In a nearby tent, Qu Zhongquan's group from Yunnan was analyzing calcium and magnesium spectra to derive coronal density and temperature. Data from dozens of wavelengths should help fill out sketchy processes in the corona and chromosphere, Qu says.

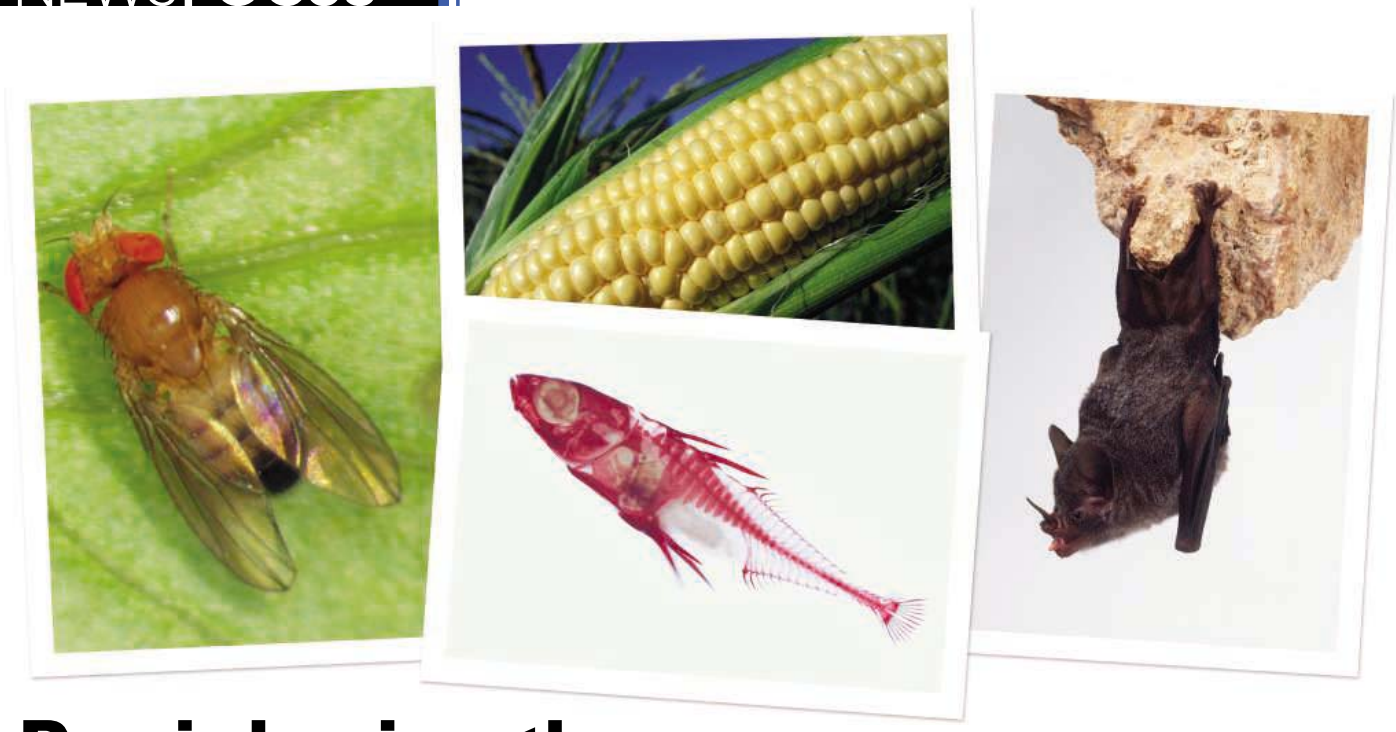
While spectra were a sure bet, a team from Purple Mountain Observatory in Nanjing was chasing a long shot. Zhao Haibin and colleagues were hoping to be the first in the world to observe vulcanoids, a class of asteroids hypothesized to exist within Mercury's orbit. Zhao's group and a second stationed in Hami, 500 kilometers to the northwest, were each using a CCD camera attached to a 15-centimeter telescope with a large field of view to image space between Mercury and the sun. Spots observed to move on complementary sets of images would be candidate vulcanoids. "The chances are small. We're going to have to get lucky," says Zhao.

Alphonse Sterling had good fortune on a different quest. Outside the Chinese scientific compound, strictly off-limits to outsiders (*Science* was granted access), the NASA solar physicist and two colleagues from the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts—Samaiyah Farid and Antonia Savcheva—were snapping high-resolution photos of coronal plumes: plasma streams that can extend several solar widths from the sun. Hinode and other satellites have obtained sharp views of plumes in extreme ultraviolet and soft x-ray spectral bands. "We want to see what these bad boys look like in white light," says Sterling, NASA's point man on Hinode. His images in Jinta were looking good, with several plumes clearly visible. Sterling's team plans to line these up with satellite data to look for differences in plumes perceived in other wavelengths.

For those who weren't lucky this time, there's always next year. A corridor cutting across the heart of China will experience a much longer total eclipse than last week's—nearly 6 minutes in Shanghai—on 22 July 2009. Chinese solar physicists have already picked out a perch near Hangzhou.

–RICHARD STONE

CREDITS (TOP TO BOTTOM): DAVID GRAY/REUTERS/LANDOV; R. STONE/*SCIENCE*



Deciphering the Genetics of Evolution

Powerful personalities lock horns over how the genome changes to set the stage for evolution

Sitting quietly in the back of the seminar room, Hopi Hoekstra doesn't stand out as a rabble-rouser. But last year, this young Harvard University evolutionary geneticist struck a nerve when she teamed up with evolutionary biologist Jerry Coyne of the University of Chicago in Illinois to challenge a fashionable idea about the molecular mechanisms that underlie evolutionary change. Egos were bruised. Tempers flared. Journal clubs, coffee breaks at meetings, and blogs are still all abuzz.

For decades, the conventional wisdom has been that mutations in genes—in particular in their coding regions—provide the grist for natural selection. But some 30 years ago, a few mavericks suggested that shifts in how genes are regulated, rather than alterations in the genes themselves, were key to evolution. This idea has gained momentum in the past decade with the rise of "evo-devo" (*Science*, 4 July 1997, p. 34), a field born when developmental biologists began to take aim at evolutionary questions. They have proposed that mutations in regulatory DNA called cis elements underlie many morphological innovations—changes in body plans from bat's wings to butterfly spots—that allow evolution to proceed. The idea has gained support from evidence that DNA outside genes—at least some of which are cis-regulatory elements—can be crucial to an organism's ability to survive and thrive over the long term.

Urging caution. Harvard's Hopi Hoekstra argues that genetic changes must be adaptive to count as important in evolution.

The zeal with which some biologists have embraced this so-called cis-regulatory hypothesis rubbed Hoekstra and Coyne the wrong way. In a 2007 commentary in *Evolution*, they urged caution, arguing that the idea was far from proven. The article sparked a sharp debate, with accusations from both sides that the other was misrepresenting and misinterpreting the literature. "What really got people upset is the tone of the paper," says Gregory Wray, an evolutionary biologist at Duke University in Durham, North Carolina. A year later, fists are still flying—the latest skirmish took place in May on the *Scientific American* Web site—and several papers prompted by the debate have just been published.

Although both sides would agree that cis-regulatory changes and mutations in coding regions of genes themselves probably both play a role in evolutionary change, the debate has become so intense that the middle ground is sometimes lost. Those on the sidelines are calling for patience. "There are strong winds from both directions," says evolutionary biologist David Kingsley of Stanford University in Palo Alto, California. "There are a handful of tantalizing examples of both coding and regulatory change, but the solution will come when lots of examples are worked out and worked out fully."

The heat has fueled more careful looks at the evidence and

CREDITS (LEFT TO RIGHT): MARK REBEIZ; MIKE SHAPIRO/UNIVERSITY OF UTAH AND DAVID KINGSLEY/STANFORD UNIVERSITY; DOUG WILSON/USDA/ARS; VERA DOMINGUES/HARVARD UNIVERSITY; FRANK GREENAWAY/GETTY IMAGES

Diversity of form. Changes in regulatory DNA are implicated, but not always proven, in the evolution of morphological traits from a variety of organisms.

a push to find more examples of cis-regulatory changes behind evolutionary modifications. It has also stimulated discussions of related ideas about how evolution proceeds in a genome: the role of transcription factors, for example, and whether evolution is predictable, with certain types of changes being caused by mutations within genes and others by alterations in nearby DNA. "I think we are on the threshold of a very exciting time," says Wray.

Regulation and evolution

Early suggestions that gene regulation could be important to evolution came in the 1970s from work by bacterial geneticists showing a link between gene expression and enzyme activity in bacteria. About the same time, Allan Wilson and Mary-Claire King of the University of California, Berkeley, concluded that genes and proteins of chimps and humans are so similar that our bipedal, hairless existence must be the product of changes in when, where, and to what degree those genes and proteins come into play. They had drawn similar conclusions from studies of other mammals, as well as birds and frogs. But the tools to track down the molecular controls on gene expression and protein production didn't yet exist.

More than 2 decades later, David Stern, a Princeton University evolutionary biologist, was probing the genetic changes that result in hairless fruit fly larvae. Typically, *Drosophila melanogaster* larvae are covered with microscopic cuticular hairs called trichomes, but not those of a relative called *D. sechellia*. In 2000, Stern found that mutations in genes were not involved and that changes in the regulation of a gene called *shavenbaby* were the cause. Sean Carroll of the University of Wisconsin (UW), Madison, saw a similar pattern in his group's studies of pigmentation patterns in fruit flies and in 2005 wrote an influential paper in *PLoS Biology* that helped convince the field that cis-regulatory changes were central to morphological evolution.

Carroll argued that mutations in cis regions were a way to soft-pedal evolutionary change. Genes involved in establishing body plans and patterns have such a broad reach—affecting a variety of tissues at multiple stages of development—that mutations in their coding regions can be catastrophic. In contrast, changes in cis elements, several of which typically work in concert to control a particular gene's activity, are likely to have a much more limited effect. Each element serves as a docking site for a particular transcription factor, some of which stimulate gene expression and others inhibit it. This modularity makes possible an infinite number of cis-element combinations that finely tune gene activity in time, space, and degree, and any one sequence change is unlikely to be broadly disruptive.

Data have been accumulating that suggest such regulatory changes are important in evolution. Take sticklebacks. In this fish, marine species have body armor and spines, but freshwater species don't. Four years ago, researchers tracked some of the difference to altered expression patterns in a gene called *Pitx1* but found no coding differences in the *Pitx1* gene of the two species (*Science*, 18 June 2004, p. 1736). "There's no doubt there's been a regulatory change," says Carroll.

Carroll, his postdoc Benjamin Prud'homme, and their colleagues discovered that closely related fruit flies vary in the pattern of wing spots used in courtship, and they have traced these changes to the regulation of a gene called *yellow* at the sites of the spots. Multiple cis-element changes—adding a few bases or losing others—have caused spots to disappear and reappear as *Drosophila* evolved and diversified, they reported in the 20 April 2006 issue of *Nature*.

Similarly, Carroll's group reported in the 7 March issue of *Cell* that various alterations in a cis element controlling a *Drosophila* gene called *tan*—which plays a role in pigmentation and vision—underlie the loss of abdominal stripes in a fruit fly called *D. santomea*. This species diverged from a dark sister species once it settled onto an island off the west coast of Africa less than 500,000 years ago.

"I'm distressed that Sean Carroll is preaching … that we know how evolution works based on such thin evidence."

—JERRY COYNE, UNIVERSITY OF CHICAGO

"I am not trying to say that regulatory sequence is the most important thing in evolution." But for morphological changes, "it's a shutout" in favor of cis elements.

—SEAN CARROLL, UNIVERSITY OF WISCONSIN, MADISON

Bat wings, too, may have arisen in part from a change in a cis element regulating a gene, *Prx1*, involved in limb elongation. Chris Cretekos, now at Idaho State University, Pocatello, and Richard Behringer of the University of Texas M. D. Anderson Cancer Center in Houston isolated this cis element in the short-tailed fruit bat and then substituted it for the mouse version of this regulatory DNA in developing mice. The resulting mice had a different expression pattern of the gene and longer forelimbs than usual, Cretekos, Behringer, and their colleagues reported in January in *Genes and Development*. The mouse and bat Prx1 protein differs by just two amino acids, which don't seem to affect its function, they note.

And there are several cases in plants where cis elements have proved important. Teosinte, the ancestor of domesticated corn, sends up multiple stalks, whereas corn grows via a single prominent one. In 2006, John Doebley and his colleagues at UW Madison linked this change to a difference in DNA several thousand bases from a gene called *teosinte branched 1*, indicating a role for noncoding cis elements in the evolution of corn.

"When you think about the sort of evolution we're interested in—why is a dog different from a fish—that has to depend on changes in gene regulation," insists Eric Davidson, a developmental biologist at the California Institute of Technology in Pasadena.

Where's the beef?

But Hoekstra and Coyne say this enthusiasm doesn't rest on solid evidence. In their *Evolution* article, they picked apart these examples

CREDITS (TOP TO BOTTOM): PAUL MERIDETH; STEVE PADDOCK

Fruit fly fashions. Mutations in regulatory DNA help explain species differences, such as abdominal stripes and no stripes (*left*) and wings with and without spots (*above*).

and the rationale behind them. They pulled quotes from Carroll's work to criticize his fervor and berated the evo-devo community for charging full speed ahead with the cis-regulatory hypothesis. "Evo devo's enthusiasm for *cis*-regulatory changes is unfounded and premature," they wrote. Changes in gene regulation are important, says Hoekstra, but they are not necessarily caused by mutations in cis elements. "They do not have one case where it's really nailed down," she says.

Coyne and Hoekstra accept only cases in which a mutation in a cis element has been demonstrated to modify a particular trait, not just to be correlated with a difference. That's "the big challenge," says Hoekstra. In the stickleback case, for example, the fact that the marine species expresses *Pitx1* where spines develop and the lake species does not—although both have the same unmodified gene—doesn't prove that a cis element is responsible for the difference, Hoekstra and Coyne argue. Even Kingsley, who works on this gene in sticklebacks, agrees that the case isn't airtight. "We still need to find the particular sequence changes responsible for the loss of *Pitx1* expression," he says.

Furthermore, the duo insist that the modified trait must be shown to be beneficial in the long run. Thus, they dismiss the *shavenbaby* example not only because causative changes in cis-regulatory elements haven't yet been identified but also because no one really knows whether the fine hairs on fruit fly larvae confer a selective advantage. "I'm distressed that Sean Carroll is preaching to the general public that we know how evolution works based on such thin evidence," Coyne told *Science*.

Coyne and Hoekstra also take issue with the notion that morphological changes are unlikely to be caused by mutations in the genes for body plans because those genes play such broad and crucial roles. Similar constraints apply across all genes, they argue. Processes such as gene and genome duplication and alternative splicing can provide room for evolutionary changes by enabling genes to take on new roles while still doing their original jobs, they note.

They point instead to a large body of evidence indicating that so-called structural changes in protein-coding genes play a central role in evolution. They list 35 examples of such changes—including a mutation in a transcription factor—in a variety of species to bolster their case. They also point out that the small differences between the chimp and human genomes, which led Wilson and King to question whether mutations in coding regions can account for the differences between the species, still add up to plenty of meaningful gene changes—an estimated 60,000. "Adaptation and speciation probably proceed through a combination of *cis*-regulatory and structural mutations, with a substantial contribution of the latter," they wrote.

Mice camouflage. Changes in the coding regions of genes underlie the coat color differences between a light, beach-dwelling subspecies of mouse and the brown mainland one.

Beyond the debate

Almost as soon as their article appeared, lines were drawn and rebuttals planned. Carroll thought he was misrepresented. "I am not trying to say that regulatory sequence is the most important thing in evolution," he told *Science*. But when it comes to what's known about the genetic underpinnings of morphological evolution, "it's a shutout" in favor of cis elements, he asserts. By not accepting that body-plan genes are a special case, Hoekstra and Coyne "muddied clear distinctions that are based on good and growing data," he charges. Carroll also doesn't buy into the requirement that the new form needs to be shown to result in a selective advantage.

Günter Wagner, an evolutionary developmental biologist at Yale University, is also critical. "There clearly are well-worked-out examples where microevolutionary changes

CREDITS (TOP TO BOTTOM): MARK REBEIZ; NICOLAS GOMPEL AND BENJAMIN PRUD'HOMME; SHAWN P. CAREY/MIGRATION PRODUCTIONS; HOPI E. HOEKSTRA/HARVARD UNIVERSITY

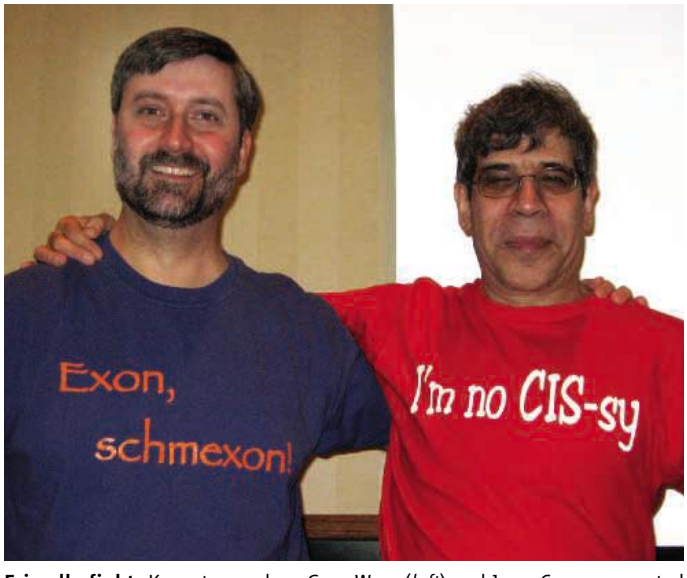
can be traced back to cis-regulatory changes," he says. Coyne and Hoekstra were "too harsh." Other evolutionary biologists grumbled that because the article was an invited perspective it didn't undergo official peer review.

On the other hand, William Cresko of the University of Oregon, Eugene, thinks it was high time for a reality check. Some researchers, he said, had become "complacent about the data." Katie Peichel of the Fred Hutchinson Cancer Research Center in Seattle, Washington, agrees: The cis-regulatory hypothesis got "taken up without [researchers] realizing there are nuances. We haven't solved morphological evolution."

In spite of the intense rhetoric, the debate has had at least some humorous moments. At the IGERT Symposium on Evolution, Development, and Genomics in Eugene, Oregon, in April, Wray—who concluded in a March 2007 *Nature Reviews Genetics* piece that cis regulation was, for certain genes, more important than structural changes—and Coyne shared center stage as the keynote speakers. Coyne's title was "Give me just one cis-regulatory mutation and I'll shut up," and he wore a T-shirt that said "I'm no CIS-sy." Wray's T-shirt said "Exon, schmexon!" suggesting that coding regions, or exons, didn't matter all that much. (Carroll couldn't make it to the meeting.) Yet in May, Carroll and "I'm no CIS-sy" faced off online on the *Scientific American* comments page.

On the positive side, the dispute has stimulated some new research. Rather than ask which type of change is more important, for example, Wray is examining whether there are any patterns in the types of mutations that are associated with different types of genes. He has scanned the human, chimp, and macaque genomes for regions that are positively selected in each species, looking for stretches conserved in two of the species but much changed in the third. He kept track of whether the region is coding or noncoding and determined which genes are involved. This computer study gives a sense of what kinds of mutations are important in the evolution of various types of genes but does not tie specific sequence changes to particular altered traits. At the IGERT meeting, he reported that genes related to immune responses and basic cell signaling have evolved primarily through mutations in coding regions. In contrast, changes in noncoding, regulatory DNA predominated for genes important for development and metabolism.

Stern has gone a step further. After looking at Hoekstra and Coyne's paper, he and Virginie Orgogozo of the Université Pierre et Marie Curie in Paris did a comprehensive literature survey to ferret out any evolutionarily important mutations, dividing them according to whether they affected physiology (building muscle cells or mediating nerve cell transmissions, for example) or morphology—affecting body plan development. Unlike Hoekstra and Coyne, they included data on domesticated species and didn't demand that the change be clearly adaptive. Overall, cis-regulatory changes represented 22% of the 331 mutations cataloged. However, in comparisons between species, cis-regulatory mutations caused about 75% of the morphological evolution, they report in an article in press in *Evolution*. The data indicate that both types of changes affect both types of traits, with cis-regulatory ones being more likely for morphological trait changes between species, Stern says.



Friendly fight. Keynote speakers Greg Wray (*left*) and Jerry Coyne promoted their take on the genetic basis of evolution with custom T-shirts.

CREDIT: COURTESY OF JOHN POSTLETHWAIT/UNIVERSITY OF OREGON

Yet even these data are inconclusive, Stern warns. Because developmental biologists focus on expression patterns, and physiologists on the proteins themselves, the former tend to find regulatory changes and the latter, coding-region alterations, potentially biasing which trait depends on which type of mutation.

Also, coding changes are more likely to be identified than changes in regulatory regions in part because once a gene is linked to a trait it is easy to assay for mutations there. "It's like shooting fish in a barrel," says Carroll. In contrast, regulatory DNA is harder to pin down. It can be close to or far from the gene itself, and a given gene could have several regulatory elements, any one of which might have the causal mutation. Thus the numbers may be misleading, a point also made by Hoekstra and Coyne. "It's really difficult to say that one's going to be more important than the other," says Stern. But it's clear that cis regulation is important, he adds. "I really want to emphasize that evo-devo [researchers] haven't come to this way of thinking simply through storytelling. We came to it through the data."

To complicate matters further, mutations in coding regions can themselves alter gene regulation. As part of their take on the debate, Wagner and Yale colleague Vincent Lynch make the case in an article published online on 22 May in *Trends in Ecology & Evolution* that mutations in transcription factors can lead to evolutionarily relevant modifications in gene expression. For example, variations in a repetitive region of the gene *Alx-4*—which codes for a transcription factor important for toe development—can alter expression patterns and change body plan in dogs. Great Pyrenees are missing 17 amino acids in this region compared with other dog breeds, and these 45-kilogram pooches have an extra toe that other breeds lack. "This is an important part of gene regulatory evolution," says Wagner.

Researchers are also trying to figure out where noncoding RNAs fit in, how gene duplications make way for change, and what roles even transposons and other repetitive DNA may play. "The important question is about finding out whether there are principles that will allow us to predict the most likely paths of change for a specific trait or situation," says Patricia Wittkopp of the University of Michigan, Ann Arbor.

With so much unknown, "we don't want to spend our time bickering," says Wray. He and others worry that Hoekstra, Coyne, and Carroll have taken too hard a line and backed themselves into opposite corners. Coyne doesn't seem to mind the fuss, but Hoekstra is more circumspect about their *Evolution* paper. "I stand by the science absolutely," she says. "But if I did it over again, I would probably tone down the language."

–ELIZABETH PENNISI

MOLECULAR BIOLOGY

Industrial-Style Screening Meets Academic Biology

A $100-million-a-year-effort to find chemicals for exploring cellular processes and drug discovery is about to move into production; skeptics say it is struggling to meet its goals

Parasitologist David Williams has spent his career studying *Schistosoma*, a type of snail-borne worm that kills 280,000 people a year in the tropics and leaves millions more with chronic liver and intestinal problems. By 2005, he had found a possible target for a drug—an enzyme the parasite requires for survival. But he had no easy way to find a molecule that would block it. Then he learned that the U.S. National Institutes of Health (NIH) was inviting researchers to submit material to be tested against a huge number of chemicals to find "hits," or biological interactions. Williams applied, was accepted, and last April, he and collaborators published the results in *Nature Medicine*: After screening 71,000 compounds, they found one, Compound 9, that inhibits the enzyme and killed at least 90% of the worms in schistosome-infected mice.

Williams is now seeking funds to develop it as a drug. "It would be pretty exciting if we could get something that would be effective for schistosomiasis," a disease whose devastation he first witnessed as a Peace Corps volunteer in Ghana, he says. The worm is beginning to show resistance to the existing drug, and a better drug is needed.

The schistosomiasis story has been touted as one of the first successes of a costly, controversial NIH program announced 5 years ago called the Molecular Libraries Initiative (MLI). It aims to bring so-called high-throughput screening, once reserved for big pharmaceutical companies, to academic scientists. Its specific goals are to develop probes for exploring cell function—small molecules that bind to protein targets—and to help find treatments for diseases that don't interest big pharma. NIH says the program, now ending a 5-year, $385 million pilot stage, has begun to pay off. Ten screening centers have produced more than 60 research probes, including a few potential drug leads. This month, NIH will move into full-scale production with grants to three large centers.

The libraries project also has a side benefit, proponents say: It has spurred scores of universities to set up their own small-molecule screening facilities (see sidebar, p. 766). "Virtually every major medical school in the country" is jumping aboard in some way, says pharmacologist Bryan Roth of the University of North Carolina (UNC), Chapel Hill.

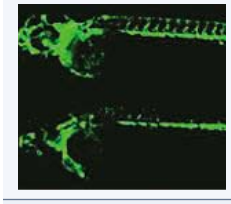
Yet even boosters of MLI acknowledge that this more than $100-million-per-year program is still an experiment—and still struggling. The screening centers took longer than expected to set up, and some were more successful than others. MLI leaders have had trouble defining certain goals, such as how strongly a compound must bind to its target to work well as a probe. NIH's original plan for sharing results has also faltered.

As the program expands, the research community remains deeply divided about it. Believers say it is generating a valuable trove of shared data and bringing rigor to the hunt for new medicines and biochemical probes. The skeptics, including several prominent drug industry leaders, aren't convinced this is a wise use of NIH's tight budget. Some worry that it may be too diffuse. It may be "a worthwhile thing to do," says Steven Paul, executive vice president for science and technology at Eli Lilly and Co. in Indianapolis, Indiana. But he asks: "Is it realistic, and is it cost effective? How potent and selective are these probes?" The answers may not become clear, some say, until nearly a billion dollars has been spent.

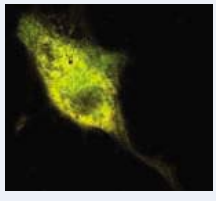
EARLY SCREENING DISCOVERIES

MEASLES
A compound that inhibits a polymerase used by the virus.

SCHISTOSOMIASIS
A chemical that killed 90% of worms in mice infected with this disease.



ANGIOGENESIS
A new inhibitor of blood-vessel formation found by screening zebrafish embryos.



GAUCHER DISEASE
Compounds that restore the ability of mutant glucocerebrosidase to process lipids in patients' cells.

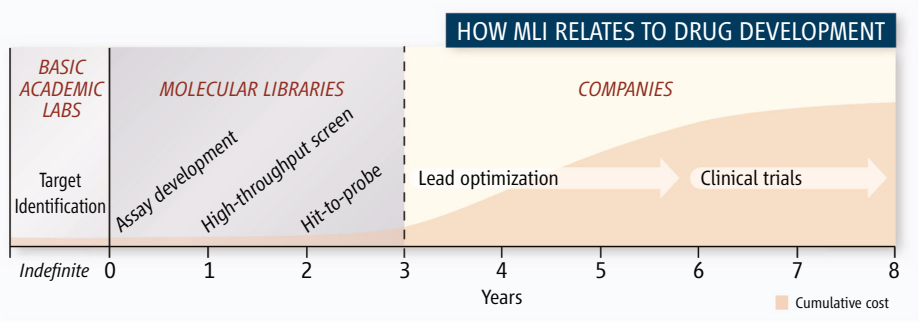
Promise. NIH's molecular screening program has produced research probes and potential drug leads for several rare or neglected diseases.

Networking

Inside a nondescript building off a busy road in Rockville, Maryland's, biotech corridor, neurogeneticist Christopher Austin presides over the NIH Chemical Genomics Center (NCGC)—a 50–staff member intramural version of the 3-year pilot screening centers NIH funded at nine external sites. At its heart is a quiet room in which three state-of-the-art yellow robots are hard at work processing biological assays. They fetch plates that are each dotted with 1536 tiny wells of different small organic molecules, mix in a protein or cell solution, then run the plate through a detector that spots whether any of the chemicals on the plates has triggered some change in the protein or cells. In another room, medicinal chemists tweak these "hits" to improve the strength and specificity of the interaction.

Although drug companies have long relied on such high-throughput screening, "this is not a world that most academic [biologists] have been in," says Austin, a former Merck researcher who says he often feels like a John the Baptist, bringing small-molecule screening to academia. The time is right for this evangelism, say Austin and other NIH officials. The explosion in genomics launched by the Human Genome Project has revealed a wealth of proteins whose functions are unknown. Some are involved in disease processes. Advances in robotics have brought down costs, making it feasible for university labs to screen a protein against hundreds of thousands of compounds, looking for one that interacts with it. That compound could then be developed into a probe that researchers would use to disrupt a protein's action or explore a cell pathway. Some, such as the schistosomiasis project, might also generate new drug leads for a tiny fraction of the overall cost of drug development (see timeline).

CREDITS (TOP TO BOTTOM): CDC; DAVID BYRES, FCCJ; RAY DINGLEDINE; O. GOKER-ALPAN/NHGRI



Filling a gap. NIH says that research probes developed through its Molecular Libraries Initiative could help fill the pipeline of potential drug leads, boosting research in early stages when costs are low.

In 2004, leaders described their plan to set up a huge central repository of 500,000 compounds that all centers would use for such screening (*Science*, 12 November 2004, p. 1138). They said that any biologist could propose screening a candidate protein, cell-based test, or even a novel assay based on a whole organism. The assay would then be peer reviewed and, if accepted, assigned to a screening center. Compounds that bind to the protein or modulate cell activity would be chemically modified until potent enough to work in a test tube but not necessarily in animals. The resulting probes would be "made available without encumbrance to all researchers," that is, without intellectual-property restrictions, Austin and other NIH leaders wrote.

MLI debuted in 2003 as the largest piece of NIH's Roadmap, a set of cross-institute initiatives. Some researchers argued that such top-down projects siphon funds from investigator-initiated science. But NIH Director Elias Zerhouni described it as a boost for basic research in his 2004 budget request to Congress, saying it would "help accelerate researchers' ability to prove the function of the complex biological circuits … in normal function and disease."

The start-up was slow. Equipping 10 academic centers to screen molecules entailed "a huge learning curve," acknowledges Carson Loomis, MLI program co-director. Initially, NIH hoped the scale-up would be similar to creating the first genome sequencing centers, he says. But high-throughput screening is not as straightforward. Centers wrestled with balky robotics equipment and chemicals that degraded. They soon realized that most of the biological assays would require many modifications to work properly when screened. They also faced the challenge of merging two cultures—biologists and chemists—and getting them to work together on a product, not hypothesis-driven research. "That interface is not a smooth one automatically," says Ray Dingledine, director of the center at Emory University in Atlanta, Georgia, and chair of the screening network.

Another challenge has been creating the small-molecule repository itself. NIH deliberately chose a wider range of chemicals than would be standard in the drug industry to make sure nothing was overlooked. But many proved "worthless" in the screens, and the ones that panned out turned out to be pretty similar to what industry would have chosen, says Christopher Lipinski, a former Pfizer chemist renowned for his skill in predicting what works as an oral drug. NIH's Linda Brady, who helped launch MLI, says the repository is growing and has improved—"I haven't heard [the term] 'junk' in a long time," she says.

One continuing debate centers on how to define an acceptable "research probe." NIH wanted the probes to be potent and selective enough to work in vitro—but no more developed than that—so that MLI participants would feel comfortable sharing raw data and forgoing patents. "There's lots of debate about where that bar ought to be," says medicinal chemist R. Kiplin Guy of St. Jude Children's Research Hospital in Memphis, Tennessee. NIH ended up loosening its original cutoffs for potency and selectivity; now it's largely up to the center to decide when a probe is complete. That has resulted in variable quality and made some centers appear more productive than others, says one center director.

On a mission. Christopher Austin, leader of NIH's screening center, hopes academics will discover the value of small molecules.

Despite the bumps, the 10-center pilot network has screened nearly 200 biological assays (far short of the projected 400) and produced 62 probes. Among these are the schistosomiasis compound; a potential drug lead for treating Gaucher disease, a rare metabolic disorder; a molecule for exploring potassium channel receptors; and probes that have shed light on the function of a new estrogen receptor. "Every center has produced at least a couple of interesting compounds," says Brady, although three—the intramural NCGC (which began a year earlier), the Scripps Research Institute's branch in Florida, and the Burnham Institute for Medical Research in San Diego, California—have produced the majority.

Missing bridges

NIH's plan for informing the broader community about these probes hasn't worked as well, however. MLI screeners must deposit screening results in PubChem, a database created as part of MLI. But these raw data reports aren't easy to use and often contain mistakes because the data aren't curated, Lipinksi says. NIH initially asked centers to post online "probe reports," Loomis says, but took them down when journal editors complained that they were too similar to submitted papers. NIH plans to require centers to post reports after a 6-month delay.

In the meantime, at *Science*'s request, NIH produced its first-ever table of completed probes. Both the total number and details of this list drew a lukewarm response from two industry experts. Some of them look "very good," says Stephen Frye, a medicinal chemist who left GlaxoSmithKline (GSK) last year for UNC, such as a measles virus inhibitor and probes for studying SP1 receptors, which are involved in sepsis. Others, however, are not very potent, he noted. Alan Palkowitz, head of medicinal chemistry at Eli Lilly, says that, based on their structures, he believes up to

CREDIT (BOTTOM): NHGRI

Universities Join the Screening Bandwagon

Once shunned as too costly and industrial, high-throughput screening is becoming a hot activity at universities. An international directory put together by the Society for Biomolecular Sciences lists 55 academic molecular screening centers—some large, some small—often paid for by a university's own budget as part of a drug-discovery program.

Unlike the screening centers funded by the U.S. National Institutes of Health (NIH) (see main text), many of these facilities lack chemists to do the tweaking required to verify a "hit"—an interaction between a chemical and a protein target—and improve the strength and specificity of the interaction. Only a few schools even have a medicinal chemistry department, says Christopher Lipinski, a retired Pfizer chemist.

Some observers say this weakness shows up in talks and papers from the new screening programs. There's a "blind spot" in academia, says Edward Spack of SRI International in Menlo Park, California: "They'll get a hit, but then many can't optimize it." Ross Stein estimates that more than 10% of the hits he sees reported in journals are false positives. "There's a lot of junk in the literature," says Stein, director of drug discovery at the Harvard NeuroDiscovery Center.

Even if academics come up with a potential therapeutic molecule, a big unknown is who will take it forward. With pharma laying off employees, and venture capital for biotechs drying up, a drug lead may have to get through preclinical animal studies before a company will pick it up, says Stein. At Merck, "a whole building of people" worked on that, says neurogeneticist Christopher Austin, a former Merck staffer who heads NIH's intramural screening center. Universities have no equivalent.

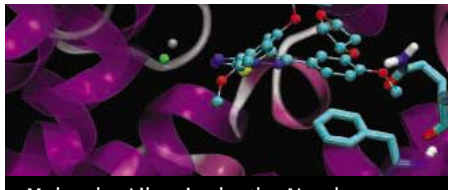
But would-be drug developers in academia note that, as part of a new push for translational research, NIH, the Wellcome Trust in the U.K., and other foundations are giving investigators money to contract out steps such as animal testing and medicinal chemistry. "If the target is important and the molecule is important, we will find a way to move it along," says molecular pharmacologist David Scheinberg of Memorial Sloan-Kettering Cancer Center in New York City.

Despite the gaps, small-molecule screening in academia is here to stay, say supporters of the approach. But there will be a shakeout. "People will either learn and get better, or they will not survive," says pharmacologist P. Jeffrey Conn of Vanderbilt University in Nashville, Tennessee. **–J.K.**

one-third of the probes might reflect spurious activity in the screens or be problematic for other reasons. He sees mostly "potential starting points" for useful probes. At the same time, both praised the list of 200-some submitted assays as including some innovative contributions such as zebrafish and tests of signaling pathways.

Arguments about quality aside, the true test of MLI will be if the broader community orders probes and starts publishing papers using them, notes Paul of Eli Lilly. However, that test may not come soon. Researchers may not have ready access to the compounds, which are often not available off the shelf. NIH is relying on center investigators to provide small amounts to the community but is not yet tracking requests in a systematic way, says Loomis. He adds, however, that a growing number of citations suggests that some probes are being used widely.

Some industry leaders question whether this massive effort is worth the time and money. If the goal is to study gene function, there are easier ways, says Peter Kim, president of Merck Research Laboratories, such as using RNAi to block gene expression and monoclonal antibodies to inhibit proteins. Small molecules are best for testing in vivo hypotheses that can lead to potential therapies, he and others say. For this, the probes usually need to be optimized to function in animals. But MLI doesn't plan to fund in vivo studies. And, says Peter Schultz of Scripps in San Diego, if academics try to do it on their own, they may face the need for the extensive medicinal chemistry and pharmacology of drug discovery. "I don't want to say the community has been swindled, but [creating selective in vivo agents is] a lot harder than it appears," says Schultz, who also oversees drug discovery as head of the Genomics Institute of the Novartis Research Foundation. (He is not involved with the Florida screening center.)



Molecular Libraries by the Numbers

Cost to date	$385 million
Pilot screening centers	10
Compound collection	~300,000
Accepted assays	268
Assays screened	191
Probes	62

Note: Data are for FY 2004–2008.

First fruits. About $385 million spent for pilot screening centers, a compound repository, a database, and technology has yielded 62 molecular probes.

MLI's leaders are used to defending against such criticism. They say small molecules are uniquely useful because they modulate the target protein directly, rather than through its gene, and can have subtle effects. "It's critical to have tools that act at the level that Mother Nature does," says Austin.

Growing investment

Despite the skepticism, reviewers who examined MLI in early 2006 concluded that it showed enough promise to continue. A project this ambitious may need 10 years to prove itself, says chemist Catherine Peishoff of GSK. "To say it's a success or failure would be unfair at this point," she says.

This month, NIH will move into what it calls "full-scale production" by funding three "comprehensive" centers for up to 6 years that will each screen 25 assays a year and have larger staffs of chemists to improve the hits. (NIH also plans to work with chemical vendors to make the probes available.) The top contenders for full-scale awards appear to be the intramural center; Scripps of Florida; Burnham; and the Broad Institute at Harvard University, which until now has had separate NIH funding for high-throughput screening. A handful of smaller centers will work on specialized screens or chemistry.

It may be expensive and risky, but MLI is important because many drug companies are abandoning high-throughput screening and shedding chemists, argues Frye, whose division at GSK was dissolved in 2007. "If the NIH doesn't pull this off, I think it's a big step backwards for drug discovery," he says.

Guy says its value will become clear over time: "It's true that people are relearning a lot of lessons," but now the data will be formally tested and widely shared. Guy says that, like the Human Genome Project, the results will be a vast expansion in public knowledge about biological systems, including targets that companies wouldn't touch before.

–JOCELYN KAISER

CREDITS: (IMAGE) COURTESY OF CRAIG THOMAS; (SOURCE) NIH CONGRESSIONAL BUDGET JUSTIFICATION FY 2004–FY 2009, MLI

On the edge. Vaquitas are vanishing quickly, due to gillnets that entangle them.

CREDIT: COURTESY OF ALEJANDRO ROBLES

CONSERVATION BIOLOGY

Can the Vaquita Be Saved?

Scientists are embarking on a last-ditch effort to help the world's most endangered marine mammal avoid the fate of its Chinese cousin, the baiji

On a recent June morning aboard the *Koipai Yu-Xa*, a research vessel plying the Gulf of California near San Felipe, Mexico, marine biologist Barbara Taylor let out a whoop of joy. The cruise was the first shakedown test of a special acoustic device, the T-POD, developed by an engineer in England. And he had just sent Taylor an e-mail message with the news she was most hoping for: "… the T-POD is full of lovely porpoise data." That meant the shy vaquitas, or Gulf of California harbor porpoises, still swim in enough numbers to be found via their calls. The cruise is part of a new effort to save the smallest cetacean from the fate of its Chinese cousin, the baiji: extinction.

The baiji (*Lipotes vexillifer*) was the first cetacean to succumb to human pressures, and many fear that the vaquita, *Phocoena sinus*, will soon be number two. "We had no idea if we would detect any vaquitas," because only an estimated 150 remain, says Taylor of the Southwest Fisheries Science Center in San Diego, California, chief U.S. scientist on the Vaquita Expedition 2008.

Scientists and the Mexican government are working on the animals' behalf. Late this summer, Mexico will launch a plan to restrict the use of gillnets that kill the vaquitas. Then in October, a full-tilt 2-month international scientific expedition will gather baseline data for what researchers hope will be the porpoises' eventual recovery. "The situation is dire," says Taylor. "The vaquita has only a few years left before it goes the way of the baiji."

The 1.5-meter-long vaquita has been known to science only since 1958, when three skulls were found on a beach. The porpoises—whose markings look like "mascara and lipstick," says Taylor—live solely in the northernmost part of the gulf. At the time of discovery, they are thought to have numbered in the low thousands. But every year, 20 to 30 vaquitas get caught in gillnets and drown. Heavy fishing and trawling for everything from shrimp to shark has sent them into a perilous decline, says Armando Jaramillo-Legorreta, a marine biologist at Mexico's Instituto Nacional de Ecología in Ensenada, who with Taylor and other scientists authored a 2007 study in *Conservation Biology* calling for "immediate action, not more data." Numerous efforts are already under way by a bevy of environmental organizations, and the vaquita is on both the U.S. and Mexican endangered species lists as well as the International Union for Conservation of Nature's Red List of species in critical danger of extinction. Small portions of its watery home are protected as Biosphere and Vaquita Reserves. But so far, nothing has worked.

Scientists last surveyed the vaquitas in 1997, counting 567. Using a model that tracks birth and death rates and fishing activity, Jaramillo-Legorreta and other scientists came up with the current estimate of 150. Because the little porpoises are difficult to spot in the murky waters they frequent, the best way to find them is with acoustic devices, says Jaramillo-Legorreta, who is also chief acoustical operator on the expedition.

Scientists "need to determine if the vaquitas' numbers are increasing or decreasing. To even be able to say that would be a major scientific advance," says Taylor, noting that the vaquitas are so timid that they have only been seen at long distances (a sighting at 900 meters is considered close) and never captured alive.

"Fortunately, we're now seeing the best efforts ever from the Mexican government to save the vaquita," says Jaramillo-Legorreta. The new plan, the Action Program for the Conservation of the Species Vaquita, was hammered out with fishers over the past 4 years. It calls for buying out boats and helping fishers start new businesses such as ecotourism, replacing gillnets with other gear, or compensating fishers for staying out of prime vaquita territory. The government is already pressing ahead, having allocated nearly $20 million for the purpose, says Luis Fueyo, the coordinator of the vaquita program in Mexico's Ministry of the Environment in Mexico City. "There are 750 licensed fishing boats" in the three main towns near the reserve, says Fueyo, "and we've purchased 308 licenses, representing 247 boats." Another 52 fishers are switching their gillnets; the remaining 451 boats will not fish in 1200 square kilometers of core vaquita habitat. There are, of course, illegal fishers in the Gulf, making law enforcement a top priority, says Fueyo.

"The nets have to come out of the water—forever," says Taylor. "That is the only way to save the vaquita. It is a huge challenge and an enforcement nightmare, but [it's] the only way." The vaquitas' maximum population growth rate is assumed to be like that of other porpoises, only 4% a year, she says. With the vaquitas' numbers so abysmally low, any growth "will be hard to detect," even with the sophisticated upcoming survey, says Taylor. She adds that the numbers may also be "politically difficult. It's hard for politicians to say that six additional vaquitas a year is good news." But for Taylor and the other Vaquita Expedition scientists, some of whom listened in vain for the sound of a baiji in 2006, six new vaquitas would be worth many a whoop of joy.

–VIRGINIA MORELL

LETTERS

edited by Jennifer Sills

The Drugs Not Taken

A MISSING ELEMENT IN CURRENT CONCEPTIONS of personalized medicine, including the Policy Forum by S. H. Katsanis *et al.* ("A case study of personalized medicine," 4 April, p. 53), is data on the patient's adherence to prescribed drug dosing regimens (*1*). Explicit data on patients' drug dosing histories are essential for sound judgments about the sources of variability in drug responses. Harter and Peck pioneered the modeling of variability in drug response, estimating that variations in patient adherence to the prescribed drug dosing regimen were a leading source of variability (*2*).

Burnier *et al*. found that nonadherence, identified by electronic monitoring of patients' drug dosing histories, accounted for about half of the nonresponders to three-drug therapy in a consecutive series of hypertensive patients (*3*). Once motivated to take the prescribed drugs, most became normotensive, with some having postural hypotension because triple therapy, taken as prescribed, was overdosing. The findings of Burnier *et al*. in conjunction with current information on the prevalence of nonadherence in all fields of ambulatory pharmacotherapy (*4*), including hypertension (5), are important to the rational implementation of personalized medicine.

JOHN URQUHART

AARDEX Group, Untermueli 6, 6302 Zug, Switzerland, and Department of Biopharmaceutical Sciences, Center for Drug Development Sciences, University of California San Francisco/University of California Washington Center, Washington, DC 20036, USA. E-mail: urquhart@ix.netcom.com

References and Notes

1. J. Urquhart, *Br. J. Clin. Pharmacol*. **54**, 212 (2002).
2. J. G. Harter, C. C. Peck, *Ann. N.Y. Acad. Sci*. **618**, 563 (1991).
3. M. Burnier *et al*., *J. Hypertens*. **19**, 335 (2001).
4. L. Osterberg, T. Blaschke, *N. Engl. J. Med*. **353**, 487 (2005).
5. B. Vrijens *et al*., *Br. Med. J*. **336**,1114 (2008).
6. Conflict of Interest: I am Chief Scientist of AARDEX Ltd, pioneer developer of electronic medication event monitors, for compiling drug dosing histories in ambulatory patients.

The Potential of Genotyping

IN THE POLICY FORUM "A CASE STUDY OF personalized medicine" (4 April, p. 53), S. H. Katsanis *et al*. propose that pharmacogenetic tests should be subject to more oversight and not be sold directly to consumers. I completely agree; moreover, physicians should order them.

However, I believe that their critique of cytochrome P450 (CYP) genotyping limitations is misguided. They selected as an example the wrong psychiatric drugs, the selective serotonin reuptake inhibitors (SSRIs). SSRIs exhibit no linear relationship between dosage and plasma concentration; wide ranges between therapeutic and toxic doses; and powerful CYP inhibition from some SSRIs (*1*). Evidence-based medicine reviews forget the importance of the pathophysiological approach in medicine (*2*). They also overlook the limitations of pharmaceutically funded randomized clinical trials, which provide an average dose for an average patient during acute treatment, whereas physicians prescribe medications for chronic illness in non-average patients (*3*).

Very limited CYP genotyping research has been conducted due to the lack of interest of pharmaceutical companies and grant agencies. Moreover, this research is not easy because pharmacogenetic testing should be used in the context of other information, such as patient characteristics (gender and age) and environment (co-medications) (*4*). Well-trained physicians are needed to implement personalized medicine (*1*).

To dismiss CYP genotyping due to lack of evidence incurs risk for those patients that need this approach most. My clinical experience (*5*), supported by pharmacological literature (*5*, *6*), indicates that CYP testing may benefit some patients using some psychiatric drugs (not SSRIs). Some subjects (less than one in a thousand) lack two CYPs that metabolize most antidepressants (*7*). After identification, they can be correctly treated by relying on current pharmacological knowledge (*7*). Evidence-based medicine focuses on average patients, whereas personalized medicine focuses on unusual subjects such as these.

JOSE DE LEON

UK MHRC at Eastern State Hospital, University of Kentucky, Lexington, KY 40508, USA.

References

1. J. de Leon, *J. Clin. Psychopharm*. **27**, 241 (2007).
2. A. R. Feinstein, R. I. Horwitz, *Am. J. Med*. **103**, 529 (1997).
3. T. Hope, *J. Med. Ethics* **21**, 259 (1995).
4. J. de Leon *et al*., *Pharmacopsychiatry* **40**, 93 (2007).
5. J. de Leon, S. C. Armstrong, K. L. Cozza, *Psychosomatics* **47**, 75 (2006).
6. J. Kirchheiner *et al*., *Mol. Psych*. **9**, 442 (2004).
7. M. Johnson *et al*. *CNS Spectrums* **11**, 757 (2006).

Correcting the Record on DNA Direct

IN REFERENCE TO THE POLICY FORUM "A case study of personalized medicine (S. H. Katsanis *et al*., 4 April, p. 53), we would like to correct some inaccuracies regarding DNA Direct and comment on the misrepresentation of the findings of the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group on cytochrome

P450 (CYP450) and selective serotonin reuptake inhibitors (SSRIs).

Katsanis *et al*. suggest that DNA Direct provides CYP450 testing for SSRIs directly to consumers, rather than through a medical provider. This is inaccurate. DNA Direct does not offer interpretation of CYP450 testing for SSRIs. DNA Direct is a Web-enabled genetic consultation company staffed by board-certified genetic counselors, with medical oversight provided by an M.D. medical geneticist. All medical genetic testing is provided according to standard medical guidelines developed under the oversight of our medical director. Secure, Web-enabled interpretation and genetic consultation regarding test results are highly personalized to the patient. We advocate for consultation with a local provider if one is available for the patient, although this is not always possible given the shortage of genetics professionals. Our patients may seek consultation directly or a physician may refer a patient for services. Our most common health care provider referral is for consultation regarding CYP450 testing for tamoxifen. No patient receives testing through DNA Direct without the involvement of a health care provider.

Katsanis *et al*. inaccurately represent the recommendations of the EGAPP working group regarding CYP450 testing and SSRIs. They imply that the conclusions of EGAPP indicate that no CYP450 testing should be offered. EGAPP concluded, however, that "there is insufficient evidence to support a recommendation for or against use of CYP450 testing in adults beginning SSRI treatment" (*1*). There are indications for CYP450 testing other than SSRI use (*2*), and some individuals taking specific SSRIs may benefit from the knowledge of their CYP450 results.

ALLAN T. BOMBARD AND TRISHA BROWN

DNA Direct, San Francisco, CA 94111, USA.

References

1. EGAPP Working Group, *Gen. Med.* 9, 819 (2007).
2. U.S. Food and Drug Administration, Table of Valid Genomic Biomarkers in the Context of Approved Drug Labels, (www.fda.gov/cder/genomics/genomic_biomarkers_table.htm).

Response

IN RESPONSE TO DE LEON, WE DO NOT MEAN to "dismiss CYP genotyping" categorically. We agree that more research is needed to understand the potential benefits of CYP testing for a wide range of drugs. We selected the case of CYP testing for SSRI selection and dosing because of the availability of a recent evidence review and recommendations by the Center for Disease Control's EGAPP program (*1*). Our question was whether commercial practices are consistent with these expert findings. We hope that the issuance of additional reviews by EGAPP will further facilitate provider decision-making regarding drug selection and dosing for a number of conditions.

In response to Bombard and Brown, we acknowledge that a company-employed physician may be involved in ordering a genetic test, but the fact remains that the consumer's own health care provider need not be involved at any stage of the process. As to the claim that "DNA Direct does not offer interpretation of CYP450 testing for SSRIs," this claim is undermined by the company's own Web site, which offers a "Drug Response Panel" for 2D6, 2C9, and 2C19 for $250 to $630. The panel may be purchased directly by a consumer with a click of the mouse. The Web site allows the consumer to "check to see if our Drug Response Panel covers your medications with our FIND tool." If one types, for

example, Paxil, the FIND tool states that a "2D6 association for Paxil was found" and the "Order testing for 2D6" button appears. The FIND tool also identifies associations with 2D6, 2C9, and/or 2C19 for Luvox, Prozac, and Celexa, accompanied by the button to order the test. DNAdirect.com provides a sample personalized report (the one shown is for Tamoxifen), which includes "information on how your genes affect drug metabolism, drugs to watch out for, next steps and more" and a "drug-specific guide" that includes "how your genes affect specific drugs, alternative treatments, and more." The Web site certainly implies that a personalized report will be provided for other drugs metabolized by CYP450, and therefore would appear to offer interpretation of CYP450 testing for SSRIs, although without access to an actual personalized report it is not possible to know exactly what information is provided to the consumer. Finally, we reject the assertion that we misrepresent the conclusions of the EGAPP working group. Our paper recognizes that EGAPP limits its recommendations to CYP testing for SSRI selection and dosing, and we draw no conclusions beyond that limited context. Bombard and Brown neglect to quote the recommendation stating that "EGAPP discourages use of CYP450 testing for patients beginning SSRI treatment until further clinical trials are completed."

SARA KATSANIS, GAIL JAVITT, KATHY HUDSON*

Genetics and Public Policy Center, Berman Institute of Bioethics, The Johns Hopkins University, Washington, DC 20036, USA.

*To whom correspondence should be addressed. E-mail: khudson5@jhu.edu

Reference

1. EGAPP Working Group, *Gen. Med.* **9**, 819 (2007).

Blue Revolution Brings Risks and Rewards

E. PENNISI'S NEWS FOCUS ARTICLE "THE blue revolution, drop by drop, gene by gene" (11 April, p. 171) on agricultural drought-tolerance work overlooked the possibility that the water deficit tolerance may be directed by metabolites. With the use of a gene that produced glutamate—an amino acid implicated in signaling and homeostasis (*1*)—drought-tolerant crops were produced first in tobacco (*2*, *3*) and then in maize (*4*). Although total free amino acid concentrations in these plants doubled, glutamate did not, leading to the hypothesis that the drought tolerance was caused by signaling that sufficient water was available, which caused an increase in molecules that store water. Mungur *et al*. (*5*) showed that alterations in the levels of many hundreds of metabolites led to changes in the abundance of shoots and roots, but there were no changes in transcript profiles, indicating that the focus on transcription activating factors (*6*) instead of metabolites may be unwise.

The risks of the new "blue" technologies should be made clear. Similar to the *NF-YB1* gene (*7*), the *gdhA* gene appears to be fooling the plant cells, causing them to maintain photosynthesis through the beginning of a drought and helping them to recover quickly when the drought ends. The weakness of such a technology is that the plants become more likely to die during more prolonged dry spells (*4*). Similarly, the *ERA1* gene alteration, which causes the plant to keep stomata open longer (*8*), risks death during longer dry spells. Whereas such a risk will be worthwhile in the short dry spells found in the west and midwest

of the United States (*4*), it is not likely to help in semiarid regions, where it may not rain for an entire season. Therefore, caution should be exercised in the use of the technologies in geological drought-prone areas of Africa.

DAVID A. LIGHTFOOT

Center for Excellence, The Illinois Soybean Center, Department of Plant, Soil, and Agricultural Systems, College of Agricultural Sciences, Southern Illinois University at Carbondale, Carbondale, IL 62901, USA.

References

1. B. G. Forde, P. J. Lea, *J. Exp. Bot*. **58**, 2339 (2007).
2. R. Ameziane, K. Bernhard, D. A. Lightfoot, *Plant Soil* **221**, 47 (2000).
3. R. Mungur, A. J. Wood, D. A. Lightfoot, *Plant Growth Regul*. **50**, 231 (2006).
4. D. A. Lightfoot *et al.*, *Euphytica* **156**, 106 (2007).
5. R. Mungur, A. D. M. Glass, D. B. Goodenow, D. A. Lightfoot, *J. Biomed. Biotech*. **2**, 198 (2005).
6. K. Century, T. L. Reuber, O. J. Ratcliffe, *Plant Physiol*. **147**, 20 (2008).
7. D. E. Nelson *et al.*, *Proc. Natl. Acad. Sci. U.S.A*. **104**, 16450 (2007).
8. Z.-M. Pei, M. Ghassemian, C. M. Kwak, P. McCourt, J. I. Schroeder, *Science* **282**, 287 (1998).

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 3 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

Bad Grades for *Science* Title

READERS OF THE SCIENCESCOPE ITEM "BAD grades for U.S. science office" (20 June, p. 1577) could be forgiven for assuming that the article was about an evaluation of the White House Office of Science and Technology Policy (OSTP). It was not. No such evaluation has occurred, and the report to which the article refers does not, in fact, grade OSTP or conclude that its performance is lacking. Those inferences are the result of the article's unfortunate and misleading title. The report is an uncontroversial catalog of historical OSTP functions, essentially all of which are being performed today. The report includes no evidence that they are not being performed excellently.

JOHN H. MARBURGER III

Office of Science and Technology Policy, The White House, Washington, DC 20502, USA.

CORRECTIONS AND CLARIFICATIONS

Reports: "Strong limit on a variable proton-to-electron mass ratio from molecules in the distant universe" by M. T. Murphy *et al*. (20 June, p. 1611). The first and penultimate paragraphs quoted a laboratory limit on the drift rate of the proton-to-electron mass ratio, μ, citing T. Rosenband *et al*., *Science* **319**, 1808 (2008). In both cases, the limit should have been from S. Blatt *et al*., *Phys. Rev. Lett*. **100**, 140801 (2008): $\dot{\mu}/\mu = (+1.6 \pm 1.7) \times 10^{-15}$ year^{-1} (where $\dot{\mu}$ is μ's time derivative).

Reports: "Structural diversity of sodium" by E. Gregoryanz *et al*. (23 May, p. 1054). Several author corrections were inadvertently omitted. On page 1055, near the top of column 3, the term "tI50" should be inserted to read "50 atoms per unit cell, tI50 (Fig. 1E)…." Later in the same paragraph, the sentence beginning "On heating…" should be replaced with "The cI16, tI50, and oC120 phases all persisted up to the melting curve on heating near 118 GPa." In the next paragraph, the first mention of oP8 should read "the oP8 phase persisted on heating up to the melting curve, but…"; later in the same sentence, the β angle should be 89.20(3)°. Finally, the Fig. 2A legend should end with the added sentence "For tI19, the host unit cell is shown."

INFECTIOUS DISEASES

Plague Through History

Nils Chr. Stenseth

Plague is a disease that has played an important role in human history; indeed, the word plague has itself become an epithet for infectious disease and the eruption of pest species beyond control. Although bubonic plague (characterized by the development of swollen and painful lymph nodes) is commonly thought of as a disease of the past, plague still represents a significant public health problem, especially in Africa, Asia, and South America (*1*). Worldwide, a few thousand human cases are reported each year, with a fatality rate between 5 and 15%. The earliest recorded major plague epidemic occurred in China in 224 BCE. Plague appeared in Europe in three long-lasting pandemic waves. The earliest, the Justinian plague, killed several million people, mainly in the Byzantine Empire, during the 6th through 8th centuries. The second wave, the "Black Death," caused some 25 million deaths between the mid-14th century and its culmination in the Great Plague of London in 1665. The third pandemic started in China in the middle of the 19th century and led to 10 million deaths in India alone.

Plague and the End of Antiquity
The Pandemic of 541–750

Lester K. Little, Ed.

Cambridge University Press, Cambridge, in association with the American Academy in Rome, 2007. 380 pp. $85, £45. ISBN 9780521846394. Paper, $29.99, £15.99. ISBN 9780521718974.

Pestilential Complexities
Understanding Medieval Plague

Vivian Nutton, Ed.

Wellcome Trust Centre for the History of Medicine at UCL, London, 2008. 138 pp. $72, £35, €52. ISBN 9780854841165.

Plague is a zoonosis, a disease in which the causative agent primarily resides in wildlife species. It is now known to be caused by the bacterium *Yersinia pestis*. This bacterium has many varieties, one of which has been linked to the three pandemics (*2*). The bacterium's principal hosts are wild rodents, and typically fleas are the transmission vector between animals. Only occasionally is *Y. pestis* transmitted to domestic rodents or other animals in close contact with humans, who may then become infected and develop bubonic plague. In some instances, infected people develop the pulmonary form of the disease, which can then be transmitted from person to person by airborne respiratory droplets. When such transmission fuels an epidemic of pneumonic plague, infected individuals (if untreated) face fatality rates of 95 to 100%. This is most likely the form that dominated past pandemics.

Plague and the End of Antiquity: The Pandemic of 541–750 focuses on the Justinian plague. The editor, Lester K. Little (a historian at Smith College in Massachusetts), and 11 other authors, primarily historians, combine findings from a variety of disciplines, including history, archaeology, epidemiology, and molecular biology. They draw on written accounts recorded in Syriac, Greek, Arabic, Latin, and Old Irish as well as excavations of burial pits, abandoned villages, and aborted building projects. The book begins with historiographical and epidemiological overviews, which are followed by discussions of the course and effects of the plague's sporadic appearances in the Near East, the Byzantine Empire, and the Latin West. The final two chapters consider the ecology, evolution, and molecular history of the Justinian plague. The authors' successful integration of insights from many fields provides a thorough account of the pandemic's origins, lethality, waxings, and wanings. The book argues, quite convincingly, that this pandemic's social, economic, political, and religious effects made it a key factor in the fading of Antiquity and the beginning of the Middle Ages.

Pestilential Complexities: Understanding Medieval Plague, edited by Vivian Nutton (Wellcome Trust Centre for the History of Medicine at University College London), spotlights the Black Death. The volume brings together skeptics and supporters (drawn from the fields of history, medicine, archaeology, and molecular biology) of the proposition that the infective agent of this second pandemic was *Y. pestis*. Collectively, the introduction and six essays offer a succinct, multifaceted account of the Black Death. The volume also places the successive waves of the pandemic that broke out in the 1340s into the wider history of the plague, looking back to the Justinian pestilence and forward toward the 19th century. And here too, the authors nicely integrate insights obtained from the several disciplines represented.

I found both books very interesting, not the least in their bridging the gap C. P. Snow discussed in his 1959 Rede Lecture (*3*). *Plague and the End of Antiquity* grew out of a 2001 gathering at the American Academy in Rome; *Pestilential Complexities* stems from a conference at the Wellcome Trust Centre. The volumes give the clear impression that the participants at these meetings really had listened to and communicated with one another. As a result, the two books mark the start of an integrated, multidisciplinary approach toward plague. They show that historians recognize that knowledge of the ecology, epidemiology, and evolution of *Y. pestis* is necessary if they are to understand the effects of plague on human history. In turn, biologists interested in the ecology and epidemiology of the bacterium will gain valuable insights from historians' studies of plague in the past.

Two additional multidisciplinary volumes integrating the sciences and humanities would complement the Little and Nutton volumes. One book could focus on the third pandemic, following it from its roots in

CREDIT: HULTON ARCHIVE/GETTY IMAGES

Plague pit. As this 19th-century engraving by J. Franklin depicts, many of the victims of the Great Plague of London (1665) were hastily buried in communal graves.

The reviewer is at the Centre for Ecological and Evolutionary Synthesis, University of Oslo, Post Office Box 1066 Blindern, Oslo 0316, Norway. E-mail: n.c.stenseth@bio.uio.n

Yunnan province through its 1894 "migration" out of Hong Kong and its subsequent spread around the world (*2*). Another could consider the plague in China. Numerous data are available on *Y. pestis* plagues in China through the centuries (*4*), and much is known about how the developments and declines of Chinese dynasties were influenced by environmental conditions, including major epidemics.

The World Health Organization considers plague a reemerging disease, and it might accurately be referred to as a neglected one. It is always worth looking back in order to understand the present—and to prepare for what might be coming. *Plague and the End of Antiquity* and *Pestilential Complexities* provide an ideal historic basis for dealing with the many facets of plague today and in the future.

References and Notes

1. N. C. Stenseth, *et al.*, *PLoS Med.* **5**, e3 (2008).
2. M. Achtman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14043 (1999).
3. C. P. Snow, *The Two Cultures and the Scientific Revolution* (Cambridge Univ. Press, Cambridge, 1959).
4. See, for example, Z. Zhang *et al.*, *Integrat. Zool.* **2**, 144 (2007).

10.1126/science.1161496

NEUROPHYSIOLOGY

A Sharp Look at Stories of Smell

Stuart Firestein

About half the lectures in olfaction seem to begin with the phrase, "and even humans are able to detect 10,000 odors." That number is also where Avery Gilbert's *What the Nose Knows*, on the sense of smell, begins. But Gilbert digs a little deeper. Why is this such a nice round number? Where did it come from? How come nobody ever gives or takes credit for it? For four pages, Gilbert tracks down the number's source in an engaging tale of scientific gullibility that should become a textbook example of how spurious facts arise and become entrenched in the literature. The answer, which I won't give away here, is like one of those jokes in which the yogi who is supposed to tell you the meaning of life turns out to be just some average Joe.

The history of olfactory science is especially curious because smell has always been intertwined with the commercial and aesthetic uses of fragrances and flavors. One could make a case that perfumery, not alchemy, gave rise to the science we now call chemistry, and odor chemistry has always involved some mixing of art and science—not always in clearly demarcated ways. Smell was long considered different from our other senses, idiosyncratic in both its physiology and psychology. Curious theories of olfaction, many based on not much more than anecdote, found their way into the mainstream and remained there much longer than was appropriate.

What the Nose Knows
The Science of Scent in Everyday Life

by Avery Gilbert

Crown, New York, 2008. 304 pp. $23.95, C$27.95. ISBN 9781400082346.

For a fresh look at the science and marketing of scent, Gilbert is well positioned at the interface between olfactory science and the fragrance industry. Although a sensory psychologist who also worked within the industry for many years, he was never really part of either academic research or the corporate culture. He therefore has an outsider's view that allows him to cast a wary and critical eye everywhere. The book reveals him to be a debunker par excellence. And olfaction, both as a discipline and an industry, has needed some debunking.

After tackling the 10,000-odors myth, Gilbert scrutinizes a series of topics including the weird chemistry underlying perfume mixtures, psychological humdingers about why we can't name odors ("not enough words," which even on the surface of it seems ridiculous), and paranoia about attempts by marketers to control consumers using odors to deliver subliminal messages to the amygdala. To each of these topics, he brings some simple sense—often reversing, or at least balancing, years of accepted drivel.

In the chapter "The olfactory imagination," Gilbert turns literary critic. Whether or not you agree with his very strong opinions on literature, you will have your eyes (or is it your nose?) opened to the ways scent and fragrance permeate literary allusion. Happily, his analysis goes far beyond the clichéd madeleines of Marcel Proust and exhibits the same thorough research and thought found in his analyses of the science and marketing of olfaction.

But it would be wrong to give the impression that this is only a book of contrariness. It offers a great deal of fun as well, and every fallacy that Gilbert debunks he carefully replaces with the facts, which invariably turn out to be more interesting. For example, there are impressively detailed chapters on the variation in olfactory ability among individuals (which helps explain why I don't smell all those awful things my wife claims to perceive), what makes a smell expert, the critical role of olfaction in foods and flavors, and the psychology of olfactory perception.

The book is also full of late-20th-century cultural references that may not be evocative to anyone who didn't live through it all but are right on the mark for those who did. A chapter on scent and the movies covers the fascinating, if doomed, history of smell-a-vision and its numerous incarnations. In the required chapter on bad smells, aptly summarized as "When bad smells happen to good people," Gilbert uses an impressive list of malodor metaphors to describe the terrible things that can result from noxious odors. And the author seems never to have heard a fart joke that he didn't like (me either). The book even has an index entry for "flatus," directing readers to a detailed two-page account of relevant experiments.

My only regret is that the book could have included more science. The field of olfaction has come of age in the past two decades, even garnering a 2004 Nobel Prize for the discovery of the olfactory receptor gene family (the largest in the mammalian genome). Gilbert doesn't mention the advances in molecular biology, physiology, and genomics that have marked the field's recent history. This rush of new discoveries has served to demonstrate how mainstream olfaction really is. Not the idiosyncratic quirky sense of just a few years ago, olfactory perception arises from mechanisms involving protein receptors, second messengers, gene transcription, axon guidance, neural regeneration—the whole shebang of modern neuroscience. And, indeed, this is precisely Gilbert's overriding thesis throughout the book: olfaction is not an enigma, a waft of incomprehensibility manipulated by a priesthood of perfumers and their strange chemical incantations. The recent developments in the field offer the best evidence for a rigorous scientific approach to all of olfaction.

In spite of all the fun, *What the Nose Knows* provides a well-researched, even scholarly, compendium of olfactory facts and fallacies, woven into an enticing history of the uses and misuses of scent. Having dug through what one can imagine must have been some very moldy smelling archives, Gilbert presents a wide-ranging yet deep look at what our "noses knowses."

10.1126/science.1162145

The reviewer is at the Department of Biological Sciences at Columbia University, New York, NY 10027, USA. E-mail: sjf24@columbia.edu

SOCIOLOGY

Scientific Misconduct: Do the Punishments Fit the Crime?

What happens to researchers after a finding of misconduct?

Barbara K. Redman[1,2] and Jon F. Merz[3]*

It is commonly accepted that punishments meted out for scientific misconduct (falsification, fabrication, or plagiarism) (*1*) effectively end one's career, banishing the bad apple for violating the trust that the scientific community confers on its members (*2*, *3*). Yet, little is known about the consequences of being found guilty of misconduct. Are punishments as severe as many suspect?

We identified from public records all investigators holding terminal degrees found guilty of misconduct by the U.S. Office of Research Integrity (ORI) between January 1994 and December 2001, inclusive. In late 2003, we examined their cases, searched for publications before and after the ORI decision, and attempted to locate these people to see if the findings had caused career changes and to interview them (*4*).

In this 8-year period, ORI found that 106 individuals had committed misconduct. Of these, 43 held terminal degrees (31 Ph.D., 8 M.D., 4 M.D./Ph.D.) and were employed in a professional, faculty, or research scientist role; we omitted students and fellows, limiting our study to those who had established research careers. All but one individual worked in nonprofit research settings. Thirty-six of these scientists were found guilty of falsification or fabrication, 10 were guilty of plagiarism, and 12 were guilty of "misrepresentation." Seventeen scientists had committed only one infraction, and the remaining 26 had committed multiple breaches.

All 43 individuals were excluded from Public Health Service (PHS) advisory boards (for a mean 3.5 years), 30 were also debarred from PHS grants and contracts (mean 3.4 years), 20 were subjected to institutional oversight (mean 3.2 years), and 14 were required to retract or correct papers. Overall, these scientists received an average of 2.5 sanctions; of 94 total sanctions levied, 58% were 3-year debarments.

[1]College of Nursing, Wayne State University, Detroit, MI, 48202 USA; E-mail: ae9080@wayne.edu. [2]Center for Bioethics, University of Pennsylvania, Philadelphia, PA, 19104, USA. [3]Department of Medical Ethics, University of Pennsylvania School of Medicine, Philadelphia, PA, 19104, USA; E-mail: merz@mail.med.upenn.edu

*Author for correspondence.

There were few differences in number or duration of sanctions between those who committed fabrication and/or falsification, plagiarism, or misrepresentation. The only systematic differences observed were (i) retraction was never required after plagiarism and (ii) those who had falsified and/or fabricated data were 8.8 times ($z = 2.34$, $P = 0.019$) more likely than others to receive grant debarments and received on average 0.6 more sanctions.

Searching PubMed, we found publication data for 37 of the 43 individuals. Papers were examined to ensure correct authorship. Mean publication rate per year before the finding of scientific misconduct (dating back to each individual's first publication) was 2.1 (SD = 1.7, range 0.2 to 5.9) and after the finding 1.0 (SD = 1.2, range 0.0 to 5.6) (dating up to late 2003). This decline was significant ($t = 4.66$, $P < 0.0001$). Twelve individuals published nothing after the misconduct finding.

From publications and other public sources, we located 28 of 43 scientists. As anticipated, many had changed jobs. Twenty-three of these 28 traceable scientists worked at universities at the time of their misconduct finding, and 10 of these were still in academia at the time of the study. Eight individuals moved to industry from university or other nonprofit positions, all of whom had been found guilty of falsification or fabrication but not plagiarism or misrepresentation.

We successfully contacted 22 of the 28 scientists by phone or e-mail. Three people did not follow up with us, and 12 expressly refused; several who refused told us they simply wished to put it behind them.

Interviews were held with seven individuals, who all reported financial and personal hardship. Six hired lawyers to defend themselves; surprisingly, three reported receiving some assistance from their institutions, one with legal help and two with nonfinancial support. Several reported that they could not appeal their cases because they lacked the resources to do so. Several became physically ill and experienced major disruptions in their personal lives.

Nonetheless, most reported that they had recovered or sustained useful scientific lives after initial shocks to their reputations. Indeed, six of the seven continued to publish in the years after the ORI determination (the exception had moved to industry). Our interviewees were more productive than the other scientists, publishing on average 1.3 more papers per year after their cases were decided ($t = 2.77$, $P = 0.0045$), and they were less likely to have been excluded from federal grants and contracts (Fisher's exact test, $P = 0.019$). Thus, the picture of the consequences painted by our interviews, which shows both the hardship of punishment and the chance for redemption, is perhaps more positive than it should be.

We found that 43% of academic scientists whom we could trace remained employed in academia after being found guilty of misconduct, and overall 19 of 37 scientists (51%) found to have committed misconduct continued to publish at least an average of one paper per year after their cases were decided. Overall, the punishments we observed were related to the crimes: Acts of falsification and fabrication were punished more harshly than were acts of plagiarism.

Of course, we have only studied those found guilty of misconduct by ORI, which is the tip of the iceberg. In the shadow of the official misconduct apparatus, there are informal means for sanctioning poor conduct that never see light beyond the bounds of the laboratory, the department, the institution, or the discipline (*5*). Whether sanctions meted out across the scientific establishment are reasonable and fairly applied requires further study.

References and Notes

1. 42 Code of Federal Regulation §50.102 (2004).
2. P. Woolf, *Hastings Center Rep.* **11**(5), 9 (1981).
3. J. B. LaPidus, B. Mishkin, in *Ethics and Higher Education*, W. W. May, Ed. (Macmillan, New York, 1990), pp. 283–298.
4. This study was approved by the Institutional Review Boards at Wayne State University and the University of Pennsylvania. Informed consent was obtained verbally during phone interviews.
5. S. L. Titus, J. A. Wells, L. J. Rhoades, *Nature* **453**, 980 (2008).
6. The authors thank E. Hockman for statistical assistance, the scientists who agreed to be interviewed, and J. Karlawish and anonymous peer reviewers for comments. This work was supported in part under RFA-RM-06-002. Opinions are solely those of the authors.

10.1126/science.1158052

IMMUNOLOGY

A Breath of Aire for the Periphery

A rare cell type in peripheral lymphoid tissues may act as a safety net in eliminating autoreactive immune cells.

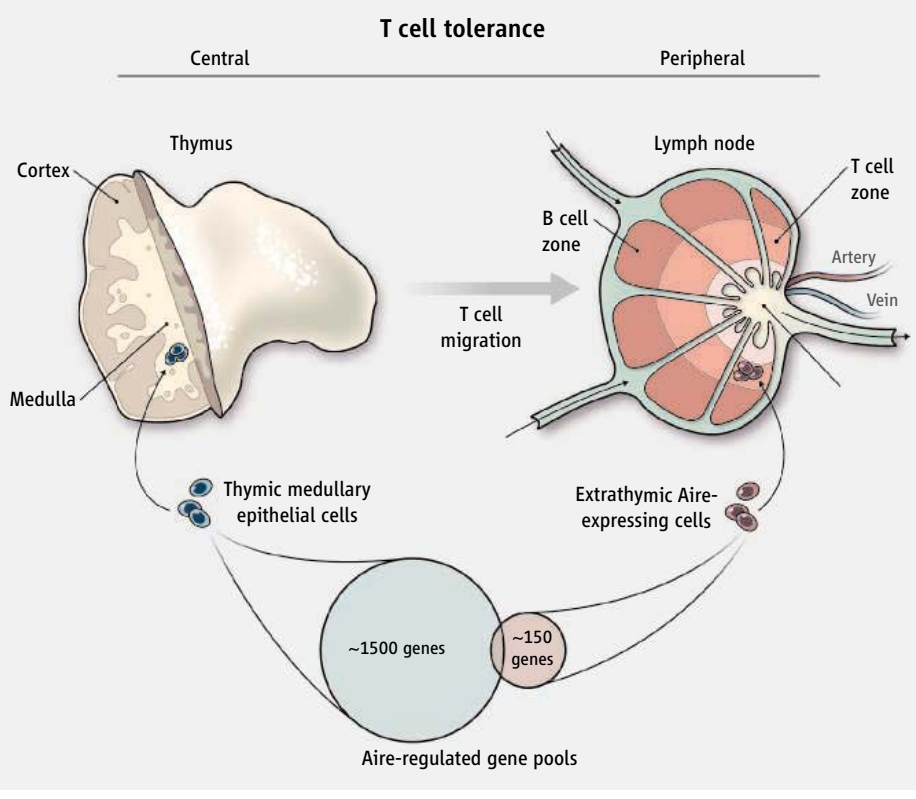
Bruno Kyewski

The term "self-tolerance" encompasses all mechanisms that protect the body against attack by its own immune system. The adaptive arm of the immune system generates immune cells that express antigen-specific receptors by a random mechanism that requires quality control—selecting a "personalized" repertoire of receptors directed against foreign but not self-antigens (*1*). Central and peripheral tolerance to self are distinguished according to the site where tolerance is imposed (*2*). Central tolerance for T lymphocytes occurs in the thymus, where their primary antigen receptor repertoire is generated. Here, developing T cells that recognize and react to self-antigens are eliminated or diverted into T regulatory cells that suppress activation of the immune system and prevent self-reactivity. Although the thymus displays a vast array of self-antigens, including those whose expression is otherwise restricted to specific tissues, this collection is nevertheless incomplete. On page 843 of this issue, Gardner *et al.* (*3*) report how peripheral lymphoid tissues act as a safety net, preventing T cells specific for antigens not presented in the thymus from escaping elimination.

Medullary thymic epithelial cells, a particular thymic stromal cell type, express a diverse set of genes that are otherwise restricted to certain tissues and/or stages of development (*4*). This so-called promiscuous gene expression in the thymus is partly regulated by a transcriptional regulator called the Autoimmune regulator (Aire). Mice deficient in Aire develop a multi-organ autoimmune syndrome, similar to that of humans with functional mutations in the *Aire* gene (*5*).

Aire is highly expressed in thymic medullary epithelial cells. However, the functionally relevant expression of Aire in cells of peripheral lymphoid organs has been controversial (*5*–*9*). Gardner *et al*. now identify cells in peripheral lymph nodes, spleen, and Peyer's patches (lymphoid structures of the gut), that express Aire and mediate deletion of autoreactive T cells. The authors genetically engineered mice in which the promoter of the *Aire* gene drives expression of a fusion protein composed of green fluorescent protein and islet-specific glucose-6-phosphatase related protein (Igrp), an antigen specific to the pancreas. Of the medullary thymic epithelial cells and peripheral cells that expressed the reporter protein, 85% and 25% expressed endogenous Aire, respectively. Most of these peripheral cells, called extrathymic Aire-expressing cells, were stromal-type epithelial cells, located at the interface between T and B cell areas in peripheral lymphoid tissues. These cells also expressed receptors characteristic of antigen-presenting cells, but differed in several markers from the medullary epithelial cells in the thymus. Surprisingly, some of these extrathymic Aire-expressing cells were highly mobile within the lymph node microenvironment, and at the same time were able to delete T cells specific for the reporter protein.

Perhaps the most intriguing result of this study relates to the target genes controlled by Aire in the thymus versus the periphery. The number of genes in the latter is about one-tenth of that in the thymus, and their degree of Aire-dependent regulation is less pronounced. Moreover, although there is little overlap between the gene pools, both are clearly enriched in genes encoding for tissue-restricted self-antigens (see the figure). The distinct composition of both gene pools favors a role for peripheral tolerance that is complementary to tolerance developed in the thymus. A recent study by Lee *et al.* also identified a fraction of non-hematopoietic cells in mesenteric lymph nodes that express Aire and certain tissue-restricted self-antigens, and mediates peripheral T cell deletion (*9*). However, these cells were less rigorously enriched, and differed phenotypically, compared to those identified by Gardner *et al.* Moreover, the data of Lee *et al.* are more in line with the concept that peripheral tolerance serves as a



Complementary tolerance. The transcriptional regulator Aire controls the expression of complementary pools of self-antigens in the thymus and peripheral lymphoid tissues that sequentially imprint central and peripheral T cell tolerance, respectively.

CREDIT: C. BICKEL/SCIENCE

Division of Developmental Immunology, Tumor Immunology Program, German Cancer Research Center, Im Neuenheimer Feld 280, Heidelberg, D-69120 Germany. E-mail: b.kyewski@dkfz.de

backup for central tolerance rather than being complementary.

The study by Gardner *et al.* still leaves some important questions that need to be answered before a definitive role can be assigned to Aire in peripheral tolerance. Why do only 25% of peripheral cells in the transgenic mice that express the fluorescent reporter protein also express endogenous Aire? Is it due to ectopic expression of the reporter construct in otherwise Aire-negative cells? Is endogenous expression of Aire too low to be detected, or is expression of Aire and the reporter protein not synchronized? The relatively low concordance between the reporter and endogenous Aire expression may also contribute to the apparently relatively low degree of gene expression induced by Aire in the peripheral cells, which for most genes is less than twofold compared to the background expression in Aire-deficient mice.

Although tolerance induction is thought to be exquisitely sensitive to low numbers of self-antigens that are presented to T cells in the context of the major histocompatibility complex, it will be essential to show that the low expression level of endogenous tissue-restricted self-antigens in peripheral Aire-expressing cells are "tolerogenic." After all, at first glance, the autoimmune phenotype of Aire-deficient mice was fully reproduced by transplanting Aire-deficient thymic stromal cells (with no functional evidence for Aire in extrathymic sites) (*5*). Given the different composition of self-antigens displayed in peripheral cells, the autoimmune phenotype caused by lack of Aire in the periphery may have been subtle and previously overlooked. Notwithstanding these caveats, the study by Gardner *et al*. raises intriguing questions about the role and function of Aire and the nature of the peripheral cells that express this factor (*10*).

The emergence of the extrathymic Aire-expressing cells in vertebrates is interesting, given that organized secondary lymphoid organs evolved much later than the thymus (*11*). *Aire* is a single-copy gene with orthologs in mammals, birds, and fish whose structure has been conserved in vertebrates over more than 400 million years (*12*). No ancestral *Aire* genes have been reported in invertebrates. This suggests that *Aire* and its role in tolerance were acquired early during vertebrate evolution, most likely concurrent with the emergence of the adaptive immune system. One question is whether Aire's only role is to ensure central tolerance, or whether it has been coopted for other functions. The study by Gardner *et al.* now presents a strong argument in favor of the latter—Aire also seems to contribute to establishing peripheral tolerance.

References

1. F. M. Burnet, *Aust. J. Sci.* **20**, 67 (1957).
2. J. Alferink, S. Aigner, R. Eibke, G. Hämmerling, B. Arnold, *Immunol. Rev.* **169**, 255 (1999).
3. J. M. Gardner *et al.*, *Science* **321**, 843 (2008).
4. B. Kyewski, L. Kein, *Annu. Rev. Immunol.* **24**, 571 (2006).
5. M. Anderson *et al.*, *Science* **298**, 1395 (2002).
6. M. Halonen *et al.*, *J. Histochem. Cytochem.* **49**, 197 (2001).
7. F.-X. Hubert *et al.*, *J. Immunol.* **180**, 3824 (2008).
8. L. A. Nichols *et al.*, *J. Immunol.* **179**, 993 (2007).
9. J.-W. Lee *et al.*, *Nat. Immunol.* **8**, 181 (2007).
10. W. W. Franke, R. Moll, *Differentiation* **36**, 145 (1987).
11. T. Boehm, C. C. Bleul, *Nat. immunol.* **8**, 131 (2007).
12. M. Saltis *et al.*, *Immunogenetics* **60**, 105 (2008).

10.1126/science.1162966

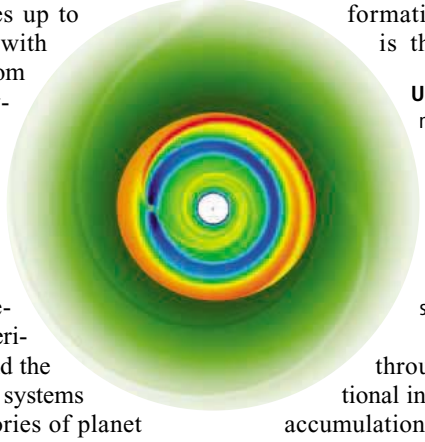
ASTRONOMY

Planetary System Formation

The diversity of extrasolar planets and planetary systems challenges present theories of planetary system formation.

J. C. B. Papaloizou

To date, 307 extrasolar planets have been discovered and 29 multiple-planet systems have been identified (*1*, *2*). The masses of the planets range from a few Earth masses up to several Jupiter masses, with orbital periods ranging from slightly over 1 day to several years. Unlike in our solar system, the orbital eccentricities of the extrasolar gas giant–sized planets may be large. On page 814 of this issue, Thommes *et al.* (*3*) describe how the range of periods, the eccentricities, and the diversity of the planetary systems have challenged the theories of planet formation, and propose an explanation—in terms of properties of the protoplanetary gas disk—of how this diversity may arise.



Under construction. Two planets, each of four Earth masses, embedded and inwardly migrating in a locally isothermal and laminar disk with uniform surface density are illustrated in the outer parts. The migration occurs because of the gravitational pull associated with the density wakes they produce. As the outer planet migrates faster, it catches up with the inner one when their orbital periods are in the ratio 3:2. These planets then migrate inward until either they enter an inner cavity close to the central star, remaining as a multiplanet system, or one of them accretes enough gas to become a giant planet, at which point it would make a gap in the disk (central region), slowing migration.

CREDIT: JOHN C. B. PAPALOIZOU

Planets are generally believed to form in a gaseous protoplanetary disk that orbits a central star in the late stages of its formation (*4*). A giant planet is thought to form either through a direct gravitational instability or through the accumulation of a solid core, which after reaching a few Earth masses undergoes rapid gas accretion (*5*,*6*). In both scenarios, formation is favored in the colder regions of the disk, at least several astronomical units away from the star.

The existence of giant planets with short orbital periods ("hot Jupiters") has led to an appreciation of the importance of orbital migration from large to small orbital radii occurring through dynamical interaction with the gaseous protoplanetary disk (*7*), during or just subsequent to the formation process (see the figure). Possible migration time scales that are well within the lifetimes of protoplanetary disks (i.e., 1 to 10 million years) suggest the importance of migration in determining the architecture of planetary systems (*8*, *9*). The gas disk is also expected to have a damping effect on the orbital eccentricity of isolated protoplanets, unless the mass is very large (*10*). A natural expla-

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WE, UK. E-mail: j.c.b.papaloizou@damtp.cam.ac.uk

nation of the high eccentricities is that they are produced by strong dynamical interactions occurring as a result of mutual gravitational interactions once the gas has been, or is in the process of being, removed (*11*).

The construction of planetary systems involves many physical processes and many bodies gravitationally interacting on long time scales. For example, the core accumulation scenario starts from the sticking together of submicrometer dust grains to produce larger particles, which then produce a swarm of planetesimals with sizes on the order of 1 km (*4*). In turn, these form the building blocks of cores large enough to accrete gas and to have a strong enough interaction with the gas disk to produce orbital migration. Many aspects of these processes, however, remain uncertain. Although individual aspects may be studied in depth by computer simulation, including all of them—together with the gravitational interactions of many planetary embryos and realistic protoplanetary disk modeling—is not yet feasible.

An ad hoc procedure has therefore been adopted for synthesizing planetary systems (*9*), which focuses on the late stages of formation and considers the evolution and gravitational interactions of a few protoplanetary bodies. Simplified prescriptions extracted from more detailed computations are used to describe how these objects accrete gas and solids from, and interact gravitationally with, the protoplanetary disk. An evolutionary model for the protoplanetary disk is an important determinant of the final outcome. At present this must be a somewhat uncertain procedure, hence the modeling has some explanatory power but does not yet have real predictive power.

Such modeling is able to produce giant planets that can migrate over the protoplanetary disk lifetime, thereby accounting for the close-in giants (hot Jupiters) when the migration stops. When several protoplanets are involved, pairs with commensurable orbital periods may be formed, several examples of which have already been observed (*10*). However, these may also involve super-Earths. For example, a system of three short-period planets having orbital periods in the approximate ratios 1:2:4 was recently announced (*12*). A possibility yet to be fully investigated is that these migrated inward, with a pair of strict 2:1 commensurabilities, until they entered a central magnetospheric cavity, a region where the disk has been expelled through interaction with the magnetic field of the central star (*13*). Later orbital evolution resulting from tidal interaction with the central star then caused the strict commensurabilities to be lost (*14*).

The number of planets that form and the amount of migration and dynamical evolution they undergo depend on the mass of the protoplanetary disk and its lifetime. Because disk lifetimes are comparable to planetary accumulation times and the latter tend to be shorter for the more massive disks, short- lived low-mass disks tend to produce few giant planets undergoing limited orbital migration, such as in our solar system. In contrast, more massive disks would produce more giant planets that undergo large-scale migration and appreciable dynamical interactions. Thus, Thommes *et al.* relate the architecture of the final planetary system to the properties of the original protoplanetary disk.

The idea investigated by Thommes *et al.* that some period of strong dynamical interactions or gravitational scattering takes place in systems containing giant planets with high orbital eccentricity is compelling. It has been shown to have reasonable success at reproducing the observed eccentricity distribution (*11*). However, it would tend to produce planetary orbits that are to some extent misaligned with the stellar equatorial plane defined by its rotation axis. Such a misalignment is now measurable through its effects on the spectrum of the central star, in systems with transiting planets. At present the indications are that several systems with close-in giant planets are consistent, with no misalignment between the orbital and stellar equatorial planes (*15*). However, the available sample is too small to either support or rule out any theory at this point. But with new planets being discovered at an accelerating pace, this situation may change in the near future.

References

1. G. W. Marcy, R. P. Butler, *Annu. Rev. Astron. Astrophys.* **36**, 57 (1998).
2. S. Udry, D. Fisher, D. Queloz, in *Protostars and Planets*, V. B. Reipurth, D. Jewitt, K. Keil, Eds. (Univ. of Arizona Press, Tucson, AZ, 2007), pp. 685–699.
3. E. W. Thommes, S. Matsumura, F. A. Rasio, *Science* **321**, 814 (2008).
4. J. C. B. Papaloizou, C. Terquem, *Rep. Prog. Phys.* **69**, 119 (2006).
5. L. Mayer, T. Quinn, J. Wadsley, J. Stadel, *Astrophys. J.* **609**, 1045 (2004).
6. P. Bodenheimer, J. B. Pollack, *Icarus* **67**, 391 (1986).
7. J. C. B. Papaloizou, R. P. Nelson, W. Kley, F. S. Masset, P. Artymowicz, in *Protostars and Planets*, V. B. Reipurth, D. Jewitt, K. Keil, Eds. (Univ. of Arizona Press, Tucson, AZ, 2007), pp. 655–668.
8. S. Udry, M. Mayor, N. C. Santos, *Astron. Astrophys.* **407**, 369 (2003).
9. Y. Alibert, O. Mousis, C. Mordasini, W. Benz, *Astrophys. J.* **625**, L57 (2005).
10. W. Kley, M. H. Lee, N. Muray, S. J. Peale, *Astron. Astrophys.* **437**, 727 (2005).
11. M. Juric, S. Tremaine, preprint available at http://arXiv.org/abs/astro-ph/0703160 (2007).
12. M. Mayor *et al.*, preprint available at http://arXiv.org/abs/0806.4587 (2008).
13. J. Bouvier, S. H. P. Alencar, T. J. Harries, C. M. Johns-Krull, M. M. Romanova, in *Protostars and Planets*, V. B. Reipurth, D. Jewitt, K. Keil, Eds. (Univ. of Arizona Press, Tucson, AZ, 2007), pp. 479–494.
14. C. Terquem, J. C. B. Papaloizou, *Astrophys J.* **654**, 1110 (2007).
15. J. N. Winn *et al.*, *Astrophys J.* **665**, L167 (2007).

10.1126/science.1161556

PSYCHOLOGY

Trust Me on This

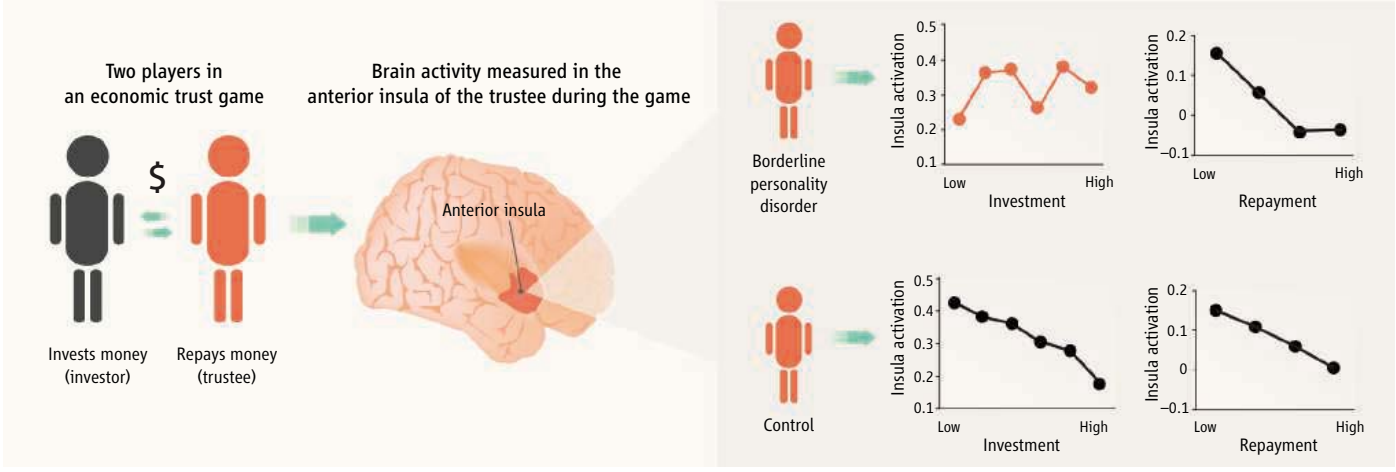
Andreas Meyer-Lindenberg

Borderline personality disorder is associated with abnormal activity in a brain region associated with monitoring trust in relationships.

Unstable interpersonal relationships, reduced impulse control, and difficulty regulating emotions characterize borderline personality disorder, a severe mental illness that accounts for up to 20% of psychiatric inpatients and exerts a tremendous toll on those afflicted, their social network, and the health-care system (*1*). Close relationships of patients are often tumultuous, spiraling out of control through highly emotional and unpredictable behavior that can leave others baffled, angry, and frightened. On page 806 in this issue, King-Casas *et al.* (*2*) use an economic exchange game and neuroimaging to provide a glimpse into the neural mechanisms underlying the breakdown of cooperation in individuals with borderline personality disorder. The study also establishes a game theory paradigm that holds promise for investigating social interactions, particularly psychiatrically relevant disturbances of social behavior.

Central Institute of Mental Health, J5, University of Heidelberg, 68159 Mannheim, Germany. E-mail: a.meyer-lindenberg@zi-mannheim.de

In the multiround economic trust game, money is exchanged between an investor, who decides how much money to commit, and a trustee, who decides how much of the investment (which is tripled during the transfer) to repay the investor. If both cooperate, both benefit from the exchange, much more so than if the investor keeps most of the money. However, this requires a degree of trust between the players, which is built up through repeated fair offers. An investor who does not trust will not invest much money. This is exactly what happened at the end of games with trustees who suffered from borderline personality disorder, indicating that they were less likely to establish or maintain a cooperative relationship. By contrast, healthy trustees were successful at doing so (thus, investment remained high at the end of the game). The better outcome was accomplished through a coaxing strategy, in which wary investors transferring small amounts of money were encouraged by generous returns, which signaled trustworthiness. Healthy players used this strategy twice as often as borderline personality disorder subjects. Why?



Just can't cooperate. Activation of the anterior insula is observed during an economic trust game in individuals with borderline personality disorder and healthy controls. Both groups show higher activation in response to stingy repayments they are about to make. However, only players with the disorder have no differential response to low offers from an investor (upper left graph), indicating that they lack the "gut feeling" that the relationship (cooperation) is in jeopardy.

To find out, King-Casas *et al*. used neuroimaging to study brain activation of trustees confronted with a small investment (a signal of the investor's lack of trust). Individuals with borderline personality disorder and healthy players differed in the activity of one brain area—the anterior insula (see the figure). In healthy trustees, small investments corresponded to large activations and large investments corresponded to small activations. By contrast, in players with borderline personality disorder, the anterior insula did not distinguish between offer sizes. As expected from previous work (*3*), the same brain area was also reactive to the amount trustees were about to repay the investor, but this was now found in both patients and healthy controls. In healthy controls, the anterior insula was activated in response both to distrustful offers from investors and stingy repayments they were about to make, whereas in trustees with the personality disorder, differential neural activity was observed only when they were repaying. Thus, their impairment selectively affected representation of the other player in the pair.

The anterior insula is traditionally associated with sensing the physiological state of the body, but strongly reacts to adverse or uncomfortable occurrences in social interactions, such as unfairness (*4*), risky choices, frustration, or impending loss of social status (*5*). This brain region also responds to the intentions and emotional state of others (*6*, *7*), and imbues them with feeling (*8*). Because rewarding aspects of social interactions have been mapped to the ventral striatum in the brain (*9*), the present results suggest that activation of the anterior insula in a social context represents a negative/aversive evaluation of perceived or planned action, perhaps associated with a feeling of discomfort. If true, this implies that individuals with borderline personality disorder may have difficulty cooperating because they lack the "gut feeling" (corresponding to the anterior insula signal) that the relationship is in jeopardy and/or expect such behavior from the outset. The correspondence of these brain findings to current psychotherapeutic practice is remarkable. The most effective treatment of borderline personality disorder (*1*), dialectical behavior therapy, is based on the assumption that patients lack skills in interpersonal self-regulation, and attempts to build these abilities.

King-Casas *et al*. interpret their observations with regard to social norms: Lack of cooperativity violates social expectations, corresponding to the insula signal, and individuals with borderline personality disorder then have atypical social norms because they fail to react to norm violations from others. This may be so. However, humans in general prefer prosocial, altruistic, fair, and trusting behaviors, which have a genetic basis (*10*). Although the neural circuitry underlying these behaviors has been studied mainly with regard to pleasure-seeking actions linked to reward (*9*), negative signals indicating lack of cooperativity or trust could also contribute to the hard-wiring of prosocial behavior in the human brain.

What causes these intriguing changes in borderline personality disorder? Despite the slim evidence, it is very likely that it arises from a combination of genetic predisposition (*11*) and severe early childhood trauma (*12*). Early traumatization has been associated with enduring dysregulation of stress responses in adults (*13*). Moreover, gene variants have been identified that modify the impact of early trauma (such as child abuse) on adult symptoms of stress (*14*). It will be of interest to determine whether such genetic variants affect insula structure and function by "genetic imaging"—that is, using neuroimaging to investigate neural mechanisms linked to genetic

CREDIT: ADAPTED BY N. KEVITIYAGALA/*SCIENCE*

variation (*15*). It will also be relevant to clarify the regulatory functions of the insula in borderline personality disorder, because this brain region functionally interacts with other limbic brain regions (such as the amygdala) that are implicated in this condition (*16*).

The use of a game theoretic approach to investigate personality disorders may be useful for studying other mental illnesses where social dysfunction is a prominent source of disability and distress, such as schizophrenia or autism. Game theory originated as an instrument of neuroeconomic analyses that assume perfect rationality of the players, and at first it came as a surprise that economic choices were in fact strongly impacted by emotional and reward-related brain processes. As King-Casas *et al.* show, it has now evolved into a tool for investigating psychopathological impairment of social interactions. Such advances are needed for patients, therapists, and researchers to grapple with social dysfunction, which is among the most impairing and least treatable components of severe illnesses such as schizophrenia.

References and Notes

1. K. Lieb, M. C. Zanarini, C. Schmahl, M. M. Linehan, M. Bohus, *Lancet* 364, 453 (2004).
2. B. King-Casas *et al.*, *Science* **321**, 806 (2008).
3. B. King-Casas *et al.*, *Science* **308**, 78 (2005).
4. A. G. Sanfey, J. K. Rilling, J. A. Aronson, L. E. Nystrom, J. D. Cohen, *Science* **300**, 1755 (2003).
5. C. F. Zink *et al.*, *Neuron* **58**, 273 (2008).
6. R. Adolphs, H. Damasio, D. Tranel, G. Cooper, A. R. Damasio, *J. Neurosci.* **20**, 2683 (2000).
7. M. Bhatt, C. F. Camerer, *Games Econ. Behav.* **52**, 424 (2005).
8. L. Carr, M. Iacoboni, M. C. Dubeau, J. C. Mazziotta, G. L. Lenzi, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5497 (2003).
9. E. Fehr, C. F. Camerer, *Trends Cogn. Sci.* **11**, 419 (2007).
10. A. Knafo, R. Plomin, *Dev. Psychol.* **42**, 771 (2006).
11. S. Torgersen *et al.*, *Compr. Psychiatry* **41**, 416 (2000).
12. M. C. Zanarini *et al.*, *Am. J. Psychiatry* **154**, 1101 (1997).
13. C. Heim, D. J. Newport, T. Mletzko, A. H. Miller, C. B. Nemeroff, *Psychoneuroendocrinology* **33**, 693 (2008).
14. E. B. Binder *et al.*, *JAMA* **299**, 1291 (2008).
15. A. Meyer-Lindenberg, D. R. Weinberger, *Nat. Rev. Neurosci.* **7**, 818 (2006).
16. C. Schmahl *et al.*, *Arch. Gen. Psychiatry* **63**, 659 (2006).

17. I thank P. Kirsch and M. Bohus (Central Institute of Mental Health) for helpful discussion.

10.1126/science.1162908

ECOLOGY

The Coming Arctic Invasion

In a future warmer climate, mollusks and other species are likely to migrate from the Pacific to the Atlantic via the Bering Strait.

Geerat J. Vermeij[1] and Peter D. Roopnarine[2]

The current episode of climate warming is having drastic consequences for animal and plant life worldwide. Besides the expected poleward expansion of temperate and tropical species and the latitudinal contraction of cold-adapted ones, an even more dramatic interoceanic invasion will ensue in the Arctic: North Pacific lineages will resume spreading through the Bering Strait into a warmer Arctic Ocean and eventually into the temperate North Atlantic.

Trans-Arctic invasion began about 3.5 million years ago during the warm mid-Pliocene epoch (*1*). A combination of northward flow through the Bering Strait, high productivity in the Bering Sea (the geographic source of trans-Arctic invaders) (*2*), favorable conditions for rapid growth and dispersal in the Arctic Ocean, and the removal through extinction of many species during the mid-Pliocene in the North Atlantic (*1*) enabled hundreds of marine lineages to colonize and enrich the biotas of the Arctic and North Atlantic. Although geochemical evidence from a core drilled near the North Pole points to perennial sea-ice cover in the Arctic beginning in the middle Miocene (14 million years ago) (*3*), the presence of mid-Pliocene temperate marine mollusks in northern Alaska and Greenland (*4*) indicates that coastal sectors of the Arctic Ocean were seasonally or perennially ice-free at that time.

Source regions of potential trans-Arctic invaders. Fifty-six molluscan lineages present in the Bering and Chukchi seas (light-blue region) have not yet participated in trans-Arctic expansion but have the potential to do so; 28 of these species extend as far north as the Pribilof Islands and Anadyrski Gulf. Another 19 mollusk species are separated from related temperate Atlantic relatives by a genetic and geographic gap. These numbers exclude North Pacific lineages whose participation in the trans-Arctic interchange during the Pliocene led to the formation of species still living in the high Arctic.

CREDIT: MAP ADAPTED FROM NASA'S EARTH OBSERVATORY (9)

In much of today's ice-bound nearshore Arctic Ocean, annual phytoplankton production is a factor of 8 to 30 lower than in the Bering Sea (*2*), with production beneath the ice accounting for 1 to 33% of annual Arctic production of phytoplankton (*5*). In the relatively ice-free mid-Pliocene Arctic Ocean, food for suspension-feeding animals would have been much more abundant, allowing many planktonically dispersing, large-bodied, fast-growing species that require high productivity to survive in that ocean and to seed populations in the temperate Atlantic. Most trans-Arctic lineages with temperate Atlantic members show genetic and geographic gaps between Pacific and Atlantic populations, indicating that post-Pliocene sea-ice expansion in the coastal Arctic Ocean ended trans-Arctic dispersals in these lineages.

[1]Department of Geology, University of California at Davis, Davis, CA 95616, USA. E-mail: vermeij@geology.ucdavis.edu [2]Department of Invertebrate Zoology and Geology, California Academy of Sciences, San Francisco, CA 94118, USA. E-mail: proopnarine@calacademy.org

Climate models and recently observed trends toward contraction and thinning of Arctic sea ice predict seasonally or perennially ice-free conditions in the nearshore Arctic Ocean by 2050 or even earlier (*6*), reestablishing a regime of temperature and productivity similar to that of the mid-Pliocene. Marine mollusks, whose past and present distributions are well documented, offer unparalleled insight into how marine species and communities are likely to respond to these future conditions.

At least 77 molluscan lineages (35% of 219 shell-bearing, shallow-water mollusk species in the northern Bering Sea) have the potential to extend to the North Atlantic via the warmer Arctic Ocean without direct human assistance (*7*). Of these, 19 have Atlantic members but are separated from them by wide geographic and genetic gaps; 2 have extinct but no living North Atlantic representatives; and 56 have not yet extended beyond the Bering Sea or the Chukchi Sea just north of Bering Strait (see the figure). The remaining 142 Bering Sea lineages are distributed throughout the Arctic and subpolar North Atlantic Oceans. The number of would-be interoceanic invaders could well be much higher, because many species with northern limits in Kamchatka and the Aleutian-Commander island arc can expand northward and therefore also become candidates for trans-Arctic invasion.

Pacific-derived species already have the largest body sizes in all ecological guilds in the Arctic and in many on the east and west sides of the North Atlantic, including mussels, barnacles, coiled grazing snails, and predatory whelks. The Bering Sea source pool contains many additional large-bodied species that may establish viable populations in the temperate North Atlantic. Given that marine invasions rarely lead to extinctions in recipient ecosystems, these trans-Arctic invaders of the future will likely enrich Atlantic biotas both by adding new lineages and by hybridizing with established species. Competitive standards in the North Atlantic will rise because of the addition of large-bodied, fast-growing species to which natives must adapt.

As in the past, few Atlantic to Pacific invasions are expected. Most of the 50 shallow-water, Atlantic-derived Arctic mollusks (out of about 180 species in the American Arctic) are small-bodied compared to both the Pacific-derived members of the Arctic fauna and to potential Pacific invaders in the Bering Sea. With the exception of the largest Atlantic-derived species (bivalves in the genera *Tridonta* and *Cyrtodaria*, with shell lengths of 38 to 50 mm), which arrived in the North Pacific when the Bering Strait first opened 5.3 to 5.4 million years ago (*8*), these species do not exceed 30 mm in maximum shell dimension and might be unable to establish populations in the Bering Sea, where competition and predation are intense.

Geographic expansions within and between oceans are generally concentrated during warm periods even in areas far from the poles. The coming warmth may therefore initiate an age of renewed interoceanic dispersal worldwide, a natural experiment that we should all anticipate with great interest.

References and Notes

1. G. J. Vermeij, *Paleobiology* **17**, 281 (1991).
2. P. K. Dayton, in *Polar Oceanography B: Chemistry, Biology, and Geology*, W. O. Smith Jr., Ed. (Academic Press, San Diego, 1990), pp. 631–685.
3. A. A. Krylov *et al.*, *Paleoceanography* **23**, PA1S06, 10.1029/2007PA001497 (2008).
4. L. D. Carter, J. Brigham-Grette, L. Marincovich Jr., V. L. Pease, J. W. Hillhouse, *Geology* **14**, 675 (1986).
5. O. G. N. Andersen, in *The Arctic Seas: Climatology, Oceanography, Geology, and Biology*, Y. Herman, Ed. (Nostrand Reinhold, New York, 1989), pp. 147–191.
6. J. Stroeve, M. M. Holland, W. Meier, T. Scambos, M. Terreze, *Geophys. Res. Lett.* **34**, L09501 (2007).
7. The values in this paragraph are from a literature synthesis, partly based on (*1*) but with updates.
8. A. Yu. Gladenkov, A. E. Oleinik, L. Marincovich Jr., K. B. Barinov, *Palaeogeogr. Palaeoecol. Palaeoclimatol.* **183**, 321 (2002).
9. R. Stöckli, E. Vermote, N. Saleous, R. Simmon, D. Herring, *Eos* **87**, 49 (2006).

10.1126/science.1160852

STRUCTURAL BIOLOGY

Symmetric Transporters for Asymmetric Transport

Nathan K. Karpowich and Da-Neng Wang

The crystal structure of a membrane transporter protein sheds light on the molecular mechanism by which glucose is absorbed by the intestine and the kidneys.

The average Western adult metabolizes hundreds of grams of carbohydrates per day, half of which is used as an energy source for the brain. To benefit from these ingested carbohydrates, they must first be broken down into simple sugars, such as glucose, and absorbed through the epithelial cells of the intestine. The glucose must then be reabsorbed in the kidneys. On page 810 of this issue, Faham *et al.* (*1*) report a major advance in elucidating the molecular mechanism by which this highly effective absorption is realized.

Glucose is absorbed from the lumen into the epithelial cell by the Na^+/glucose cotransporter SGLT1 in the intestine and by the related SGLT2 in the kidney (see the figure). The glucose is then exported by the glucose transporter GLUT2 to the basal side of the cell into the blood. These transport proteins share the same substrate and all function as so-called secondary membrane transporters, but they are members of distinct protein families: SGLT1 and SGLT2 are solute sodium symporters (SSS), whereas GLUT2 is a member of the major facilitator superfamily (MFS). Faham *et al.* now report the structure of a bacterial SSS protein.

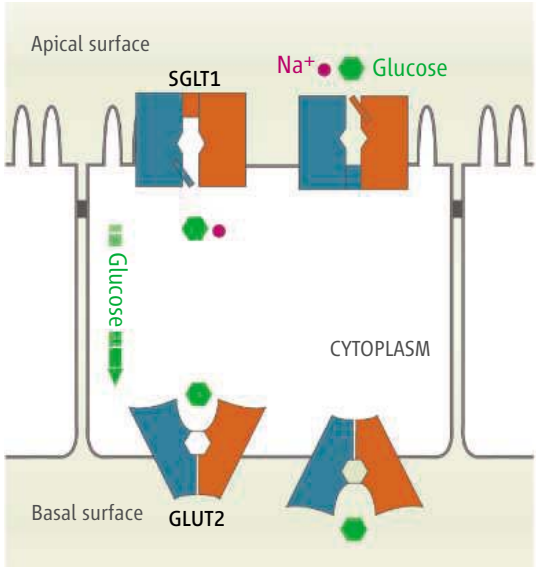
Secondary membrane transporters couple "uphill" translocation of substrate across the membrane to the energetically favorable flow of ions down their concentration gradient. Both the substrates they transport, ranging from ions and sugars to lipophilic drugs, and their protein architectures are diverse: More than 200 distinct families can be classified on the basis of primary structure. Nevertheless, biochemical, kinetic, and structural studies suggest that all secondary membrane transporters operate via a common alternating-access mechanism (*2*).

In this mechanism, the transporter is believed to have two major alternating conformations, inward-facing (C_i) and outward-facing (C_o). Interconversion between the two conformations facilitates substrate translocation across the membrane. This kinetic scheme can be realized through two types of conformational changes: a rocker-switch movement of the two halves of the protein (*3*) and an alternating gate or gated-pore mode (*4*)—that is, a channel-like protein with two gates that open alternatively (see the figure). MFS proteins are thought to

Kimmel Center for Biology and Medicine at the Skirball Institute of Biomolecular Medicine, and Department of Cell Biology, New York University School of Medicine, 540 First Avenue, New York, NY 10016, USA. E-mail: karpowic@saturn.med.nyu.edu; wang@saturn.med.nyu.edu

operate via the rocker-switch mode, and two MFS members have been visualized in their C_i conformation at high-resolution by x-ray crystallography (*5*, *6*). Structures of several other transporter proteins presumed to operate via the gated-pore mode have also been determined (*7*–*9*), but none of these structures are of proteins from the same transporter family. Hence, it remains unclear whether any transporter works as a gated pore.

Faham *et al.* now report the structure of vSGLT—an SSS protein from *Vibrio parahaemolyticus* that shares 30% sequence identity with human SGLT1 (*1*). The core of the protein is made of two inverted, symmetric halves. The substrate (galactose along with a Na^+ ion) is bound at the center of the core helices in the middle of the membrane. The substrate is closed off to the extramembrane space by a thick gate made of hydrophobic residues. The cytoplasmic gate is much thinner, consisting of only a few side chains, and a large cavity is found outside this gate. Therefore, the structure presents the substrate-bound, C_i conformation of the protein.



Gated pores and rocker switches. Net sugar transport across the intestinal epithelia is catalyzed by SGLT1 in the apical membrane, which couples Na^+ influx down its concentration gradient to glucose uptake, and by GLUT2, which facilitates glucose transport across the membrane to the basal side of the cell. A similar transport pathway is responsible for glucose reabsorption in the kidneys. Faham *et al.* show that in gated pores such as SGLT1, the rotation of two broken helices permits alternating access of the substrates to the opposing sides of the membrane. Rocker switches are thought to function by the rotation of two domains toward one another, alternately exposing the substrate-binding site to each side of the membrane.

The vSGLT structure resembles that of LeuT, a bacterial member of the neurotransmitter sodium symporter (NSS) family (*9*). The core helices of vSGLT share only 11% sequence identity with LeuT, but the two transporter families may nevertheless share a common ancestor (*10*, *11*). The LeuT crystal structure is in its substrate-bound, C_o conformation (*9*), whereas the vSGLT structure is in the C_i conformation (*1*). Faham *et al.* mapped the vSGLT sequence onto the LeuT structure and generated a C_o model for vSGLT. Comparison of the two conformations allowed the authors to propose a reasonable model for the Na^+-coupled galactose transport across the membrane by vSGLT. Such internal symmetry has recently also been used to model the C_i conformation of LeuT (*12*).

On the cytoplasmic side, the vSGLT C_i structure lacks a vestibule found between the gate and the tip of a hairpin loop on the symmetric, extracellular side of the LeuT C_o structure (*9*). This vestibule might function as a reservoir for concentrating the substrate from the extracellular space for substrate import. Recent work on LeuT has shown that the substrate, leucine, binds to a low-affinity secondary site in the vestibule and that tricyclic antidepressants inhibit leucine transport by direct competition at this site (*13*–*15*).

Both SSS-type gated pores and MFS-type rocker switches are composed of two symmetry-related homologous domains, but their interdomain association is distinct. Gated pores like SGLT are formed by the association of two V-shaped domains intertwined in an antiparallel manner. The substrate-binding site is located in the center of the membrane at two broken helices, one from each domain. The vSGLT structure shows that alternating access is achieved by rotating the cytoplasmic portion of two broken helices from the substrate-binding site. In contrast, in MFS, two parallel helical bundles associate to form a shared binding site at their interface. Substrate binding at this interface from one side of the membrane drives domain closure, thereby alternately exposing the binding site to opposite sides of the bilayer.

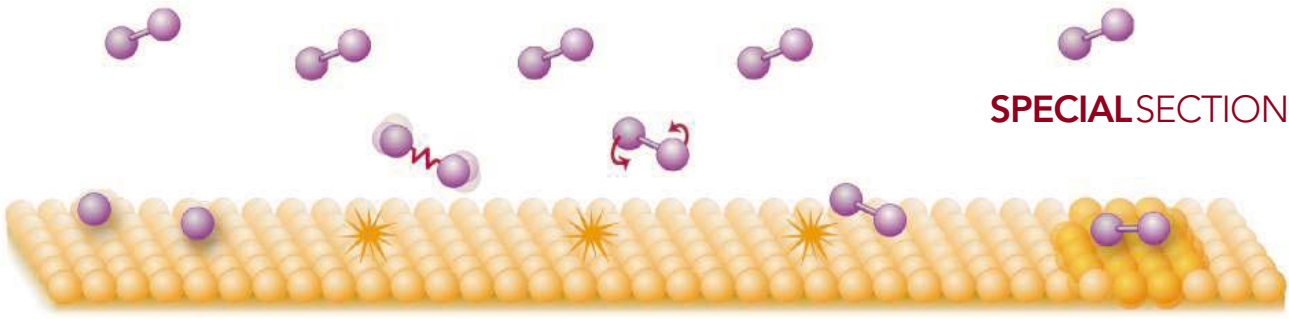
The gated-pore and rocker-switch modes both use electrochemical gradient-driven membrane transport. However, because the conformational changes required for transport in a gated pore are typically smaller than in a rocker-switch transporter, the less dynamic nature of gated pores may offer a potential advantage in crystallization experiments. The differences do not stop here. A gated pore resembles an ion channel with two gates instead of one and can often display uncoupled, channel-like ion conduction. Such uncoupled current has been observed in several NSS proteins (*16*). Likewise, in the absence of glucose, SGLT1 displays channel-like behaviors (*17*). In contrast, the single thick gate of rocker switches prevents diffusion of ions across the membrane, and MFS proteins do not display a leak current.

The vSGLT structure provides an important piece of the puzzle in understanding the molecular mechanism of sugar absorption in the human body. To completely understand this physiological process, however, there is an urgent need to identify and better characterize all conformational states in the transport cycle for a particular protein. Isolation of secondary transporters in specific conformational states remains a major technical obstacle for structural studies. However, with structural information for some conformational states in hand, spectroscopic techniques can provide precise information about the nature and magnitude of the conformational states during the transport cycle. Now with both conformational states of a gated pore visualized, the field anxiously awaits the determination of a rocker-switch structure in its outward-facing conformation.

References and Notes

1. S. Faham *et al.*, *Science* **321**, 810 (2008).
2. I. C. West, *Biochim. Biophys. Acta* **1331**, 213 (1997).
3. G. A. Vidavar, *J. Theor. Biol.* **10**, 301 (1966).
4. C. Patlak, *Bull. Math. Biophys.* **19**, 209 (1957).
5. Y. Huang, M. J. Lemieux, J. Song, M. Auer, D. N. Wang, *Science* **301**, 616 (2003).
6. J. Abramson *et al.*, *Science* **301**, 610 (2003).
7. E. Pebay-Peyroula *et al.*, *Nature* **426**, 39 (2003).
8. D. Yernool, O. Boudker, Y. Jin, E. Gouaux, *Nature* **431**, 811 (2004).
9. A. Yamashita, S. K. Singh, T. Kawate, Y. Jin, E. Gouaux, *Nature* **437**, 215 (2005).
10. J. S. Lolkema, D. J. Slotboom, *FEMS Microbiol. Rev.* **22**, 305 (1998).
11. E. Turk, E. M. Wright, *J. Membr. Biol.* **159**, 1 (1997).
12. L. R. Forrest *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10338 (2008).
13. Z. Zhou *et al.*, *Science* **317**, 1390 (2007).
14. S. K. Singh, A. Yamashita, E. Gouaux, *Nature* **448**, 952 (2007).
15. L. Shi, M. Quick, Y. Zhao, H. Weinstein, J. A. Javitch, *Mol. Cell* **30**, 667 (2008).
16. J. N. Cammack, S. V. Rakhilin, E. A. Schwartz, *Neuron* **13**, 949 (1994).
17. D. D. Loo *et al.*, *J. Physiol* **518**, 195 (1999).
18. N.K.K. is a recipient of an American Heart Association Postdoctoral Fellowship and of a Ruth L. Kirschstein National Research Service Award from NIH. The authors' work was financially supported by the NIH (DK053973, GM075936, GM075026, and MH083840).

10.1126/science.1161495



page 794

INTRODUCTION

Theoretical Possibilities

NOT SO LONG AGO, EXPERIMENTS IN THE SYNTHETIC CHEMISTRY LAB TENDED TO proceed without detailed calculations and predictions. Today, most organic chemists rely on the results of density functional theory (DFT) calculations to help decide between alternative synthetic routes and to provide guidance on structure assignments. However, such readily solved problems are not where theoretical chemists focus their attention for their own current day-to-day work—rather, it is on the problems that still defy solution. This special section highlights some areas where theorists feel that progress can be made in the near future on problems that will benefit their experimental colleagues in chemistry and other disciplines.

Stone (p. 787) describes the ubiquitous first step in tackling theoretical chemistry problems: constructing a potential energy surface, which describes the energy of the molecule or assembly of molecules as a function of the atomic positions. Clary (p. 789) goes on to detail the status of predicting reactions between isolated small molecules in the gas phase—a context in which the highest level of quantum-mechanical detail is currently attainable. Reactions involving three atoms have been modeled successfully in this way, but even four- or six-atom reactions remain challenging. Although DFT has been the workhorse of most computational studies, it is not reliable in certain classes of problems; for example, it often underestimates transition-state barriers in reactions or band gaps of materials. Cohen *et al*. (p. 792) trace the problems of many DFT approximations to delocalization and static correlation errors through a framework that makes use of fractional charges and fractional spins.

Practical chemistry often involves molecular interactions in multiple phases beyond gaseous collisions. Kroes (p. 794) discusses progress in modeling molecule/surface interactions, which play a central role in heterogeneous catalysis. Much effort has focused on unraveling the contexts in which electronic and nuclear motion become coupled in time. Klein and Shinoda (p. 798) describe simulations of complex molecular systems, in which a vast number of simultaneous interactions are encompassed through molecular dynamics trajectories. Use of so-called coarse-grain models allows the behavior of micrometer-scale systems of polymers or biopolymers to be modeled over the course of microseconds, a time scale that is highly relevant for comparison to experiments.

Theoretical chemistry is now commonly used to address complex problems in biochemistry and materials science. In a News story, Service (p. 784) describes recent successes in simulating protein folding, a problem long hindered by the computational intractability of the immense number of accessible configurations. Carter (p. 800) moves beyond molecular systems to focus on simulations of extended materials. Recent progress in ab initio and DFT methods has facilitated purely theoretical explorations of features ranging from mechanical properties to corrosion behavior. Electronic excitation remains a challenging frontier.

A recurring theme in all of these articles is the complementary role of theory and experiment in exploring chemical questions. Each approach nourishes the other, presenting fresh challenges.

– JULIA FAHRENKAMP-UPPENBRINK, PHIL SZUROMI, JAKE YESTON, ROBERT COONTZ

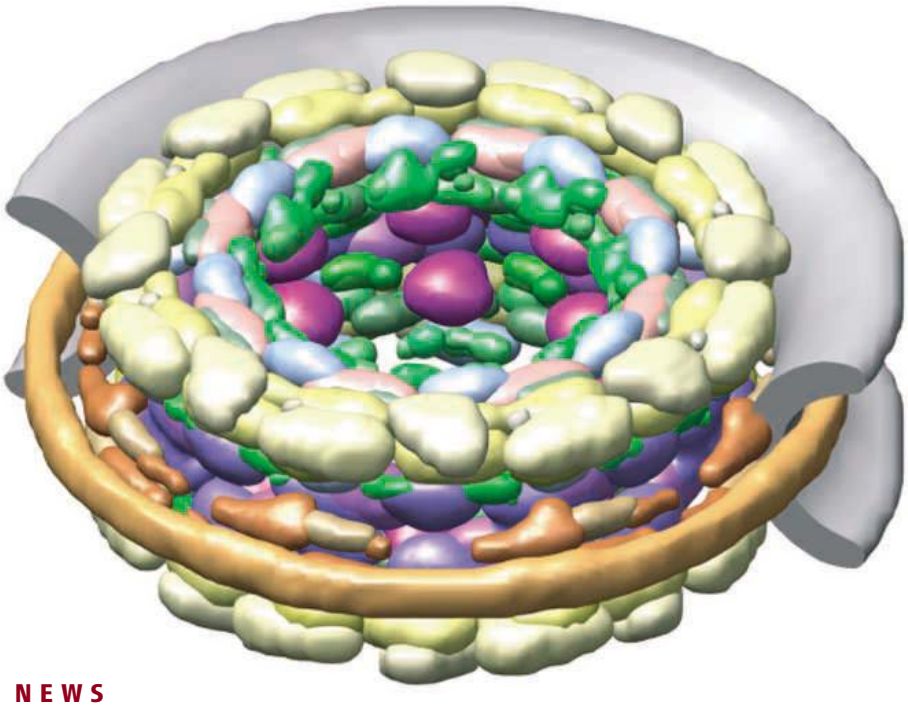
Challenges in Theoretical Chemistry

CONTENTS

News

Perspectives

Science



Big picture. Experimental data helped computer modelers nail down the structure of the nuclear pore complex.

CREDIT: F. ALBER *ET AL.*, *NATURE* (29 NOVEMBER 2007)

NEWS

Problem Solved*
(*sort of)

Researchers have toiled for decades to understand how floppy chains of amino acids fold into functional proteins. Learning many of those rules has brought them to the verge of being able to make predictions about proteins they haven't even discovered

IN 1961, CHRISTIAN ANFINSEN, A BIOCHEMIST at the U.S. National Institutes of Health, saw something that continues to perplex and inspire researchers to this day. Anfinsen was studying an RNA-chewing protein called ribonuclease (RNase). Like all proteins, RNase is made from a long string of building blocks called amino acids that fold up into a particular three-dimensional (3D) shape to give RNase its chemical abilities.

Anfinsen raised the temperature of his protein, causing it to unravel into a spaghettilike string. When he cooled it back down again, the protein automatically refolded itself into its normal 3D shape. The implication: Proteins aren't folded by some external cellular machine. Rather, the subtle chemical push and pull between amino acids tugs proteins into their 3D shapes. But how? Anfinsen's insights helped earn him a share of the 1972 Nobel Prize in chemistry—and laid the foundation for one of biology's grand challenges.

With an astronomical number of ways those chains of amino acids can potentially fold up, solving that challenge has long seemed beyond hope. But now many experts agree that key questions have been answered. Some even assert that the most daunting part of the problem—predicting the structure of unknown proteins—is now within reach, thanks to the inexorable improvements in computers and computer networks. "What was called the protein-folding problem 20 years ago is solved," says Peter Wolynes, a chemist and protein-folding expert at the University of California, San Diego.

Most researchers won't go quite that far. David Baker of the University of Washington, Seattle, believes that such notions are "dangerous" and could undermine interest in the field. But all agree that long-standing obstacles are beginning to fall. "The field has made huge progress," says Ken Dill, a biophysicist at the University of California, San Francisco (UCSF).

The work has huge implications for medicine. Misfolded proteins lie at the heart of numerous diseases, including Alzheimer's and Creutzfeldt-Jakob disease. Understanding how proteins fold could shed light on why they sometimes misfold and could suggest ways to intervene. Accurate protein models can also lead to the development of more-conventional drugs that block or enhance the work of key proteins in the body.

Twin challenges

Today, the protein-folding challenge boils down to two separate but related questions. First, what general rules govern how, and how quickly, proteins fold? Second, can researchers predict the 3D shape that an unknown protein will adopt?

These simple questions open the door to a world of mind-boggling complexity. Because two neighboring amino acids can bind to each other at any one of three different angles, a simple protein with 100 amino acids can fold in 3^{200} different ways. Somehow, a folding protein sorts through all those possibilities to find the correct, or "native," conformation.

And it's not by trial and error. Even if a folding protein could try out one different conformation every quadrillionth of a second, it would still take 10^{80} seconds—60 orders of magnitude longer than the age of the universe—to find the right solution. Because most proteins fold in milliseconds to seconds, something else is clearly going on.

Molecular biologist Cyrus Levinthal pointed out this paradox in 1969 and concluded that proteins don't follow a random set of wiggles to find their native conformation. But figuring out what path they do take hasn't been easy. Early on, researchers largely assumed that a protein follows a set path as it folds, wending its way through certain intermediate states as it coils up into its particular arrangement of helixes and sheets and so on. But in the mid-1980s, Wolynes and Dill suggested that proteins, rather than working like origami—in which one fold in a sheet of paper leads to the next until the final shape is reached—actually break the problem down into many pieces. Local clusters, each containing a handful of amino acids, initially pull

and repel one another. As these clusters begin to fold, neighboring clusters come together, and so on.

To explain this process, Wolynes and Joseph Bryngelson, both then at the University of Illinois, Urbana-Champaign (UIUC), suggested that as proteins fold they follow an energy landscape, akin to water flowing downhill. The result is a more energetically stable arrangement. Dill pushed the same notion and later came up with an image of an energy funnel, showing how proteins can follow many possible different pathways to their native conformation at the bottom of the funnel (see top figure, p. 786).

Dill's funnel explained how proteins could avoid Levinthal's paradox and fold quickly. It also led to a testable hypothesis. The time it takes for a protein to fold depends on the energetic obstacles in its path. With fewer amino acids, most small proteins fold more quickly than larger ones that can get caught up on energetic plateaus before finding another downhill run. If the simpler-is-faster rule were true, researchers realized, then it should be possible to make some proteins fold faster by mutating the amino acids that slowed things down.

In 2007, Martin Gruebele, a chemist at UIUC, and colleagues set a record for such streamlining when they tracked the folding of both native and mutated conformations of a protein called λ repressor. After making their proteins, Gruebele's team cooled them down to unravel them and then zapped them with a laser. The nanosecond burst of heat caused the proteins to begin refolding, which the Illinois team could watch by tracking their fluorescence. Certain mutations enabled the protein to fold in just 2.5 microseconds, 200 times faster than the natural protein does.

Such mutations, however, often disrupt the protein's chemical function. The reason, Gruebele says, is that in most proteins, hydrophobic amino acids tend to shield themselves from interacting with water—an energetically favorable arrangement—by nestling in the center. Charged, or polar, amino acids by contrast tend to stick out into the water that surrounds proteins. These groups tend to be more chemically active and commonly play key roles in the protein's reactive center. And as proteins wiggle into shape, the polar interactions are often slower to make their adjustments. Changing some of those amino acids speeds things up but alters the chemistry. "They evolved to do a job, not to fold fast," Gruebele says.

"What was called the protein-folding problem 20 years ago is solved."
—PETER WOLYNES, UCSD

The next step

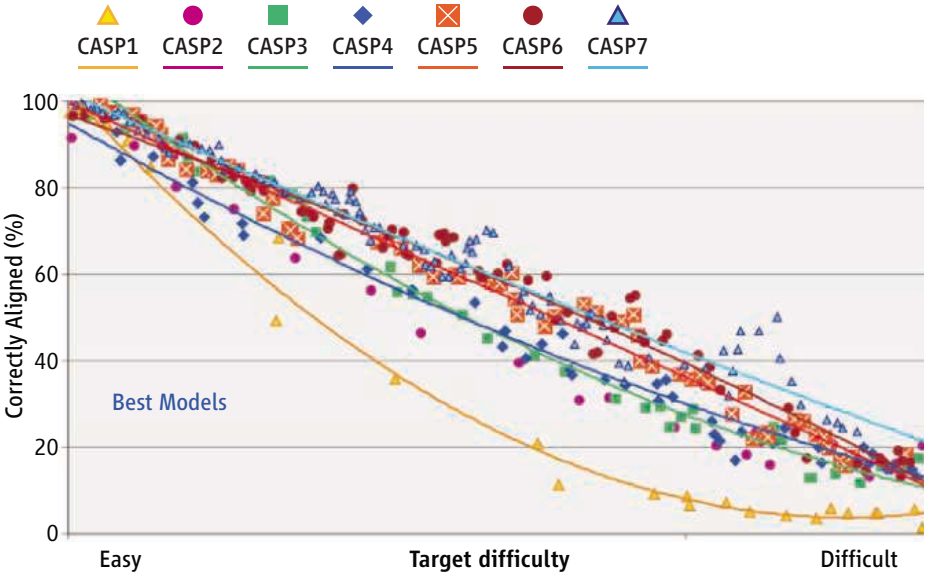
Thanks to this and many other related studies, Andrej Sali, a biochemist and protein modeler at UCSF, says that most protein-folding experts now believe they understand the general rules for how proteins fold and how they fold so quickly. But the next step—predicting how a specific set of amino acids will fold—remains a much bigger challenge. "We have not been able to transfer our conceptual understanding into [a] prediction of how specific amino acid sequences will fold," Sali says.

There have been some successes. Every 2 years since 1994, for example, computer modelers have vied to determine the 3D structure of an unknown amino acid sequence in a competition known as the Critical Assessment of Techniques for Protein Structure Prediction. At first, only about half of the modelers came close to predicting the structures of moderately difficult target proteins (see figure, below). In 2006, however, 80% did. Most of the predictions still can't match the resolution of an x-ray crystal structure, which can pinpoint the position of atoms down to a couple of tenths of a nanometer. Nevertheless, Dill says, "it's gotten to the point where if you have a reasonably small protein, you can get a good structure."

Computer models for predicting protein structures come in many varieties. But they generally fall in two camps: Ab initio models start by specifying the attractive and repulsive forces on each point and then calculate a structure by cranking through the calculations until they find the lowest energy state. Homology models, by contrast, make their predictions by comparing the target protein with proteins with closely related sequences whose structures are already known. More powerful computers and search algorithms have recently given ab initio models a major boost, but Dill says homology models still hold the upper hand.

The accuracy of those predictions depends on a model's resolution (whether it aims to map the position of individual atoms or just of individual amino acids) and how thoroughly it samples the energy landscape to find the lowest energy configuration. As a result, modelers face a tradeoff. Increase the resolution by mapping out all the atoms, and you limit the amount of sampling a computer, or network, is able to carry out. Increase the sampling rate, and you limit the resolution.

To simplify the computations, some researchers bolster their computer models with experimental data that narrow the search for the protein's lowest energy configuration. In the 25 March issue of the *Proceedings of the National Academy of Sciences*, for example, Baker and 15 colleagues in the United States and Canada described a new technique for using nuclear magnetic resonance (NMR) data to boost the speed and accuracy of protein simulations with atomic resolution (see bottom figure, p. 786).



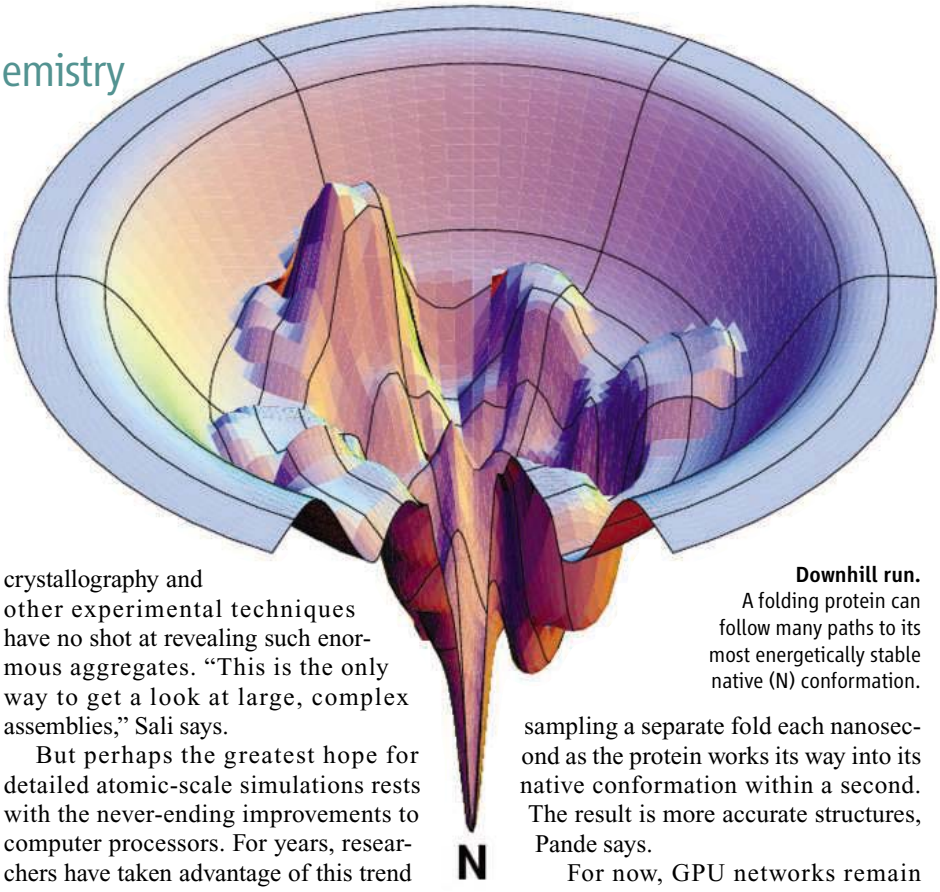
CREDIT: A. KRYSHTAFOVYCH *ET AL.*, *PROTEINS: STRUCTURE, FUNCTION, AND BIOINFORMATICS* (5 OCTOBER 2007)

Steady rise. Computer modelers have slowly but steadily improved the accuracy of the protein-folding models.

Like x-ray crystallography, NMR has long been used to map proteins in atomic detail. But the technique typically works only with small proteins. It usually requires taking at least two separate types of NMR data, an easily acquired data set known as the chemical shifts and a much slower technique called the Nuclear Overhauser effect (NOE).

In their new work, Baker and his many colleagues dispensed with NOE data and fed chemical shift data for 16 proteins into a computer prediction program known as ROSETTA. The resulting atomic-scale models closely resembled structures previously solved by either NMR or x-ray crystallography. As a control experiment, the researchers also solved the structures for nine proteins for which the NMR and x-ray structures were still being worked out. Those results, too, ended up in tight agreement. "Our joint hope is by combining our methods with NMR data we can work to larger and larger proteins," Baker says. He and his colleagues are adding other types of experimental data, such as data from lower resolution electron cryomicroscopy and from mutation experiments that highlight amino acids sitting next to each other in the folded protein.

The multipronged approach is paying off for other researchers as well. In the 29 November 2007 issue of *Nature*, Sali and colleagues in the United States, the Netherlands, and Germany predicted the structure of the nuclear pore complex, an assembly of 456 separate proteins, by integrating 10 different biophysics and proteomic data sets into a model. The resolution was low by the standards of x-ray crystallography. However, crystallography and other experimental techniques have no shot at revealing such enormous aggregates. "This is the only way to get a look at large, complex assemblies," Sali says.



Downhill run. A folding protein can follow many paths to its most energetically stable native (N) conformation.

But perhaps the greatest hope for detailed atomic-scale simulations rests with the never-ending improvements to computer processors. For years, researchers have taken advantage of this trend by joining thousands of processors together to build powerful supercomputers, such as IBM's Blue Gene machines, that have long excelled at protein-folding simulations.

More recently, Baker, Vijay Pande of Stanford University in Palo Alto, California, and other researchers have created distributed supercomputers. They rely on computer users from around the world to download software that lends the computer's central processing unit (CPU) to folding calculations when the computer is not in use. Today, Pande's Folding@home network counts more than 250,000 active participants, and Baker's Rosetta@home totals more than 300,000.

Pande says these networks have sped up protein predictions 100,000-fold. Other recent improvements to search algorithms have boosted speeds another 1000 times. And most recently, distributed networks have begun turning to ultrafast graphics processors known as GPUs to gain another 100- to 1000-fold advantage. Taken together, these improvements now allow distributed networks to follow a protein through a billion gyrations, sampling a separate fold each nanosecond as the protein works its way into its native conformation within a second. The result is more accurate structures, Pande says.

For now, GPU networks remain smaller than their CPU counterparts. Pande's GPU network, for example, counts only 10,000 participants. But these are likely to grow quickly as modelers gain experience with writing code to take advantage of their talents. "The sampling part of the problem will soon be an obsolete issue," Pande predicts.

So is the protein-folding problem solved? Not quite. But Pande and others say researchers are getting tantalizingly close. "The practical issue in protein folding is now an engineering issue," says Sali. Pande agrees: "We're on the verge of being able to tackle the complete problem."

Crossing that line won't solve all the problems in protein folding. Drug designers in particular have a tough challenge, because they often need to know the position of atoms in an active site of a protein at an ultrahigh resolution in order to design drugs to block or enhance the protein's work.

Still, Dill calls the progress in his field a revolution. But he says most scientists haven't noticed because it has occurred so slowly. "Progress in science comes out as news when there are big steps," he says. "In protein folding, there have been a huge number of folks involved and lots of incremental steps. And that doesn't usually make news."

Maybe not. But now it's set to make a difference for scientists, physicians, and their patients. **–ROBERT F. SERVICE**

Tight fit. Adding data from nuclear magnetic resonance experiments improves the accuracy of computer models of how proteins fold.

CREDITS (TOP TO BOTTOM): KEN A. DILL AND HUE SUN CHAN, *NATURE STRUCTURAL & MOLECULAR BIOLOGY*, JANUARY 1997, 4 (1); Y. SHEN *ET AL.*, *PNAS* (25 MARCH 2008)

PERSPECTIVE

Intermolecular Potentials

Anthony J. Stone

Computer simulations of molecular materials require a compact but accurate description of the energies of interacting molecules and the forces between them.

Computer simulation has become an essential aspect of research in all areas of science, from aeronautical engineering to zoology. In molecular science, the ideal simulation would involve the solution of the equations of quantum mechanics at every step to obtain the energy and possibly the forces on the system. Although a near-first-principles approach is possible by using density functional theory (DFT), it is too demanding in computer resources for most applications, and it has limitations. In particular, DFT does not recover the attractive dispersion (van der Waals) forces between molecules, although methods have been proposed in which an empirical dispersion term is added (*1*, *2*).

Accordingly, it is still necessary in many applications to adopt the well-established procedure of proposing or calculating a potential energy function that describes the energy of the system as a function of the positions of the atoms and then using this to study the motion of the system by classical mechanics (molecular dynamics, MD), to sample the energy at random configurations and average them to obtain static properties (Monte Carlo, MC), or to explore the surface (the energy landscape) to find the energy minima and the pathways between them (*3*). Even where the region of interest in a large system can be studied quantum mechanically, we often need to embed it in a reasonably realistic but simpler "molecular-mechanics" description of the rest of the system (the QM/MM approach), and a potential energy function is needed for the MM part. We are concerned here with nonbonded interactions only; potentials for describing chemical reactions are discussed in the article by D. C. Clary in this issue (*4*).

For some purposes, the exact form of the potential energy function is not critical. It is possible to learn a great deal about the general characteristics of gases, liquids, and solids by applying simulation methods with simple potentials. This was the approach used, of necessity, when computers were much less powerful than they are now, but it is still valuable in applications such as protein folding or supercooled liquids and glasses if the objective is to explore generic features rather than to predict the behavior of a particular system. Conversely, there are applications where the quality of the potential is paramount, and calculations with an inadequate potential are a waste of effort. One such application is the prediction of crystal structures of molecular solids. Most molecules of any size, including the molecules of pharmaceutical drugs, can crystallize in many different forms, or polymorphs, and it is important for commercial as well as scientific reasons to know which of many hypothetical forms, differing in lattice energy by only a few kJ mol^{-1}, are the most stable.

University Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. E-mail: ajs1@cam.ac.uk

Contributions to the Potential

A potential energy function needs to incorporate information about the system it describes, in the form of numerical parameters. Generally, the simpler forms of potential have few parameters and are fitted to a range of experimental data, such as crystal structures and sublimation energies, whereas the more elaborate potentials have too many parameters to be fitted successfully in this way and depend on calculations from first principles as well as experimental data. Much effort has been put into improving intermolecular potentials, for example by refining the simple models by fitting to a wider range of data or by introducing more elaborate and detailed functional descriptions. However, simple models are incapable of the high accuracy needed in some applications, and then it is necessary to use more elaborate functional forms and to obtain most or all of the parameters from first-principles calculation.

Except for very small molecules, it is usual to write the intermolecular potential as a sum of atom-atom terms

$$U^{AB} = \sum_{a \in A} \sum_{b \in B} u^{ab}(R_{ab}, \Omega_{ab}) \qquad (1)$$

where the sum is taken over the atoms a of molecule A and the atoms b of molecule B, and u^{ab} is some function of the distance, R^{ab}, between the atoms and possibly of their relative orientation Ω^{ab}.

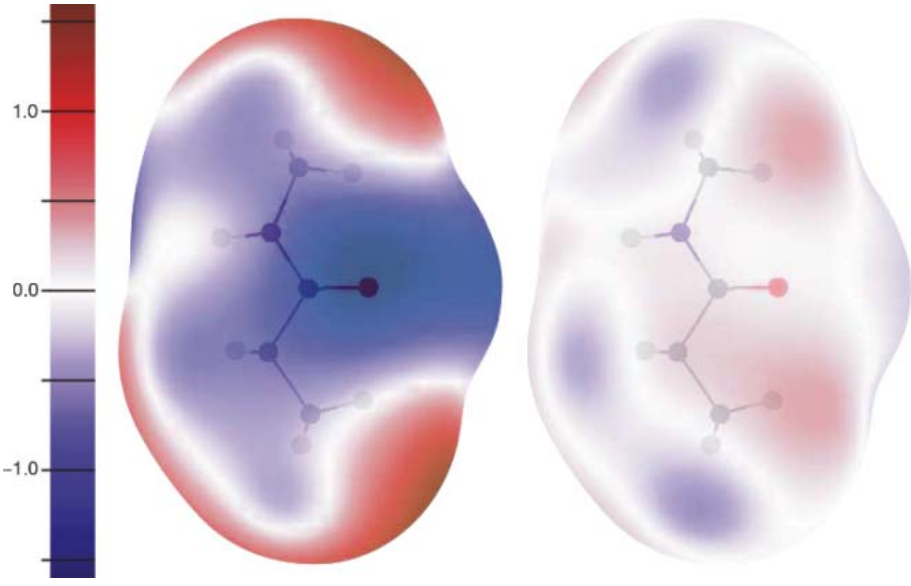


Fig. 1. Errors (V) in electrostatic potential on a surface at 1.8 times van der Waals radii around N-methyl propanamide for two models. (**Left**) Point charges; (**right**) charge, dipole, and quadrupole on C, N, and O; charge and dipole on H. The errors are much reduced in the multipole approach.

The main contributions in any of these descriptions are as follows:

1) The electrostatic term, which is the classical interaction between the charge distributions of the molecules.

2) The exchange-repulsion term (van der Waals repulsion), which arises from the overlap of the charge distributions as a consequence of the Pauli principle, which forbids two electrons with the same spin being at the same point in space simultaneously.

3) The dispersion term (van der Waals attraction), which is a nonclassical interaction arising from correlated fluctuations of the electrons in the interacting molecules.

4) Induction, where the electron density of a molecule distorts in response to the electric field of other molecules in its neighborhood.

5) Charge transfer, where incipient chemical bonding can arise from transfer of electron density from one molecule to another. Strictly speaking, this is part of the induction term, but it is sometimes treated separately.

For the electrostatic term, it has long been the practice to describe the charge distribution of each molecule by means of point charges, commonly on the atoms, but it has also been recognized for a long time that this is inadequate, especially in hydrogen-bonded systems (*5*, *6*). Distributed multipole analysis (*7*) derives a description in terms of charges, dipoles, and so on from a calculated wave function. This is a much better description than that from point charges (Fig. 1), and modern molecular dynamics code such as DL_MULTI (*8*) can handle such descriptions. A similar description can be provided by using the methods of atoms in molecules (AIM) (*9*).

Multipole descriptions account for the distortion of atoms resulting from bonding to their neighbors and for features such as π bonds and lone pairs, but they do not account for the effects of overlap of the molecular charge distributions at short range (penetration effects). It is common to absorb the resulting errors into the short-range exchange-repulsion term. An alternative approach, which is unfortunately rather expensive in computer resources, is to use Gaussian multipoles, as in the SIMPER scheme (*10*). The Gaussian multipoles can, in principle, provide an exact description of the charge distribution of the molecule.

The exchange-repulsion term is still often described by R^{-n} atom-atom terms, typically with $n = 12$ as in the Lennard-Jones potential. This gives a rather poor account of the repulsive region, and the Born-Mayer form, Ae^{-BR}, is much better. For some purposes, especially for close contacts involving halogen atoms, it is necessary to introduce some orientation dependence into the Born-Mayer form to describe the nonspherical shape of the atoms.

Dispersion (van der Waals attraction) is a universal effect. Formally, it can be described at long range by a series in the intermolecular separation, R:

$$U_{\text{disp}} = -C_6R^{-6} - C_7R^{-7} - C_8R^{-8} - \ldots \quad (2)$$

but, for all but the smallest molecules, it is necessary in practice to replace this series by a sum of atom-atom terms of similar form. The dispersion coefficients, C_n, depend on the nature of the interacting atoms; they can also depend on the relative orientation of the atoms, although this dependence is almost always ignored. Moreover, the series is usually truncated at the first term.

The dispersion series (Eq. 2) evidently diverges as the atom–atom distance tends to zero, but the dispersion energy remains finite. To deal with this, each term, C_nR^{-n}, may be multiplied by a damping function $f_n(\beta R)$, which suppresses the singularity at $R = 0$. Little is known about the form of these damping functions, and it is usual to use the functions proposed by Tang and Toennies (*11*).

The induction term is a troublesome one. It arises from the response of each molecule to the electric fields arising from the surrounding molecules. In some cases, such as inert gases or hydrocarbons, these fields are zero or very small, and the induction term can be ignored. The induction energy, in its simplest form, is $U_{\text{ind}} = -(1/2)\alpha_{ij}F_iF_j$, where F_i is the electric field due to the neighbors and α_{ij} is the polarizability tensor of the molecule under consideration. It is possible for the fields due to the neighbors to cancel out, so that the induction energy due to all neighbors may be small even though the induction energy in the presence of a single neighbor would be large. That is, the induction energy cannot be treated as a sum of pairwise atom-atom or molecule-molecule terms. Treating it properly is computationally expensive, and the usual treatment has been to approximate its effect by modifying the properties of each molecule. In water, for example, the effect of the neighbors on a particular molecule is to enhance its dipole moment, on average, so this averaged enhanced dipole moment is used and any fluctuations are ignored. It is increasingly recognized, however, that this approximation is unsatisfactory for accurate work and for highly polar systems such as proteins and, indeed, water itself.

Potential models that incorporate induction explicitly are becoming more common. For very small molecules, the overall molecular polarizability can be used, but for molecules with more than a few atoms some form of distributed treatment is needed. The simplest approach is to assign isotropic dipole-dipole polarizabilities to each atom or to selected atoms and to solve the induction problem iteratively at each step of an optimization or simulation. Including explicit induction requires complete reparametrization of the potential, because the parameters of conventional force fields without explicit induction already include induction in an average way, as explained above.

Capabilities and Limitations

The quality of current practical potentials varies inversely with the size of the molecules to be studied. Interatomic potentials for small molecules can be calculated to high accuracy by using conventional ab initio quantum mechanical methods or obtained by fitting a suitable functional form, such as that described above, to spectroscopic data. Accurate potentials for water have been obtained from both approaches (*12*, *13*).

For larger molecules, up to 20 or 30 atoms, accurate values for the parameters in a potential model of the form described above can be calculated by using symmetry-adapted perturbation theory based on DFT (*14*). Although conventional DFT cannot obtain the dispersion energy directly, coupled Kohn-Sham perturbation theory can be used to calculate polarizabilities, from which accurate dispersion coefficients can be constructed. This approach has been used to predict successfully the crystal structure of 1,3-dibromo-2-chloro-5-fluorobenzene in the 2007 blind test (*15*). The atomic multipole moments were calculated by using DFT, the atomic polarizabilities and dispersion coefficients were obtained from coupled Kohn-Sham perturbation theory, and the exchange-repulsion parameters from symmetry-adapted perturbation theory based on DFT [SAPT(DFT)]. An alternative approach, using DFT with an empirical dispersion term to calculate energies and forces directly rather than via an analytical potential energy function (*2*), was able to predict all the structures correctly in the same blind test, but it is still a much more expensive method.

For larger molecules, such as proteins, the simplest description of the nonbonded interactions uses atomic point charges, Lennard-Jones or Born-Mayer repulsions, and isotropic C_6R^{-6} atom-atom dispersion. To these are added intramolecular terms describing bond stretches, bends, and torsions. Several programs and very many parameter sets have been proposed in this framework, but there is no consensus as to which set is best, and indeed all of them have limitations, giving good results in some applications and not in others (*16*, *17*). More modern versions use distributed multipoles rather than point charges, and some incorporate induction effects explicitly (*18*, *19*), but a definitive all-purpose force field is some way in the future.

The main limitation of analytical potential functions at present is in the description of flexible molecules. It might be thought sufficient to allow the properties of each atom to move with the atom, and indeed this is the approach used by all the standard force fields for proteins. However, it is not adequate for accurate work (*20*): for example, changes in bonding during torsional motion can lead to changes in atomic charges. Describing such changes, for example by a short Taylor or Fourier series, would increase the number of parameters in the force field by an order of magnitude, and fitting the parameters to the data is likely to be a very ill-conditioned problem. At present, there is no successful general treatment of molecular flexibility.

A further limitation is the assumption of transferability. All current force fields deal with a large number of atom types: a dozen or so carbon atom types, depending on the coordination and type of substituent; a dozen or so hydrogen atom types; and so on. Increasing the number of atom types allows for a more detailed description of the systems being studied and potentially for

more accurate results, but in practice the number of atom types and hence the reliability of the calculations are limited by the quantity and quality of the experimental data used to parametrize the force field. With recent developments in the power and accuracy of ab initio methods, it will be possible to overcome this limitation and to obtain much more reliable potentials for a wide range of applications in molecular science.

References

1. T. Schwabe, S. Grimme, *Phys. Chem. Chem. Phys.* **9**, 3397 (2007).
2. M. A. Neumann, M.-A. Perrin, *J. Phys. Chem. B* **109**, 15531 (2005).
3. D. J. Wales, *Energy Landscapes* (Cambridge Univ. Press, Cambridge, 2003).
4. D. C. Clary, *Science* **321**, 789 (2008).
5. A. D. Buckingham, P. W. Fowler, *Can. J. Chem.* **63**, 2018 (1985).
6. G. M. Day, W. D. S. Motherwell, W. Jones, *Cryst. Growth Des.* **5**, 1023 (2005).
7. A. J. Stone, *J. Chem. Theory Comput.* **1**, 1128 (2005).
8. M. Leslie, *Mol. Phys.*, in press.
9. L. Joubert, P. L. A. Popelier, *Mol. Phys.* **100**, 3357 (2002).
10. R. J. Wheatley, T. C. Lillestolen, *Int. Rev. Phys. Chem.* **26**, 449 (2007).
11. K. T. Tang, J. P. Toennies, *J. Chem. Phys.* **80**, 3726 (1984).
12. Y. Scribano, N. Goldman, R. J. Saykally, C. Leforestier, *J. Am. Chem. Soc.* **110**, 5411 (2006).
13. R. Bukowski, K. Szalewicz, G. G. Groenenboom, A. van der Avoird, *Science* **315**, 1249 (2007).
14. A. J. Stone, A. J. Misquitta, *Int. Rev. Phys. Chem.* **26**, 193 (2007).
15. A. J. Misquitta, G. W. Welch, A. J. Stone, S. L. Price, *Chem. Phys. Lett.* **456**, 105 (2008).
16. M. R. Shirts, J. W. Pitera, W. C. Swope, V. S. Pande, *J. Chem. Phys.* **119**, 5740 (2003).
17. M. Feig, A. D. MacKerell, C. L. Brooks, *J. Phys. Chem. B* **107**, 2831 (2003).
18. D. P. Geerke, W. F. van Gunsteren, *J. Phys. Chem. B* **111**, 6425 (2007).
19. T. D. Rasmussen, P. Ren, J. W. Ponder, F. Jensen, *Int. J. Quantum Chem.* **107**, 1390 (2007).
20. G. Murdachaew, K. Szalewicz, *Faraday Discuss.* **118**, 121 (2001).

10.1126/science.1158006

PERSPECTIVE

Quantum Dynamics of Chemical Reactions

David C. Clary

Substantial recent progress has been made in developing theory to account for quantum effects on chemical reactions.

Theoretical and computational methods are widely used to study chemical reactions in the gas phase (*1*, *2*). Current methods enable useful and reliable calculations on the rates of chemical reactions and are also used to predict more detailed aspects of the dynamics and mechanisms of reactions that can be observed with methods such as laser spectroscopy and molecular beams (*3*).

The ease with which any such calculations can be performed depends on the topology of the potential energy surface (i.e., the energy of the electrons in a system of atoms that varies as the positions of the atoms are changed). If there is no energy barrier to reaction, very simple theories are often easily applied, and quantum dynamical effects can be minimal (*4*, *5*). Also, the method of classical molecular dynamics is widely applicable when quantum tunneling and vibrational zero point energy effects are not important (*6*).

The most challenging reactions are those for which there is an energy barrier to the reaction and in which hydrogen or deuterium atoms are explicitly involved in bond breaking and bond formation. For these reactions, quantum tunneling through the energy barrier can be a major effect, and quantum dynamics methods need to be used.

The starting point for quantum dynamics calculations is the Schrödinger equation, which needs to be solved for the electrons and nuclei taking part in the chemical reaction. Because the equation cannot be solved analytically, theories are required that allow for an efficient numerical solution. The Born-Oppenheimer approximation usually applies, allowing the electronic and nuclear motion to be treated in separate calculations. A potential energy surface is first computed through solving the Schrödinger equation for the electrons, with the atoms held in fixed positions. It is expensive computationally to calculate the electronic energies for molecular geometries that involve bonds being broken and formed in chemical reactions, and only the most sophisticated ab initio electronic structure methods yield reliable reaction barriers (*7*, *8*).

Even if the required electronic structure calculations can be carried out accurately, it is not a trivial task to fit the electronic energies to

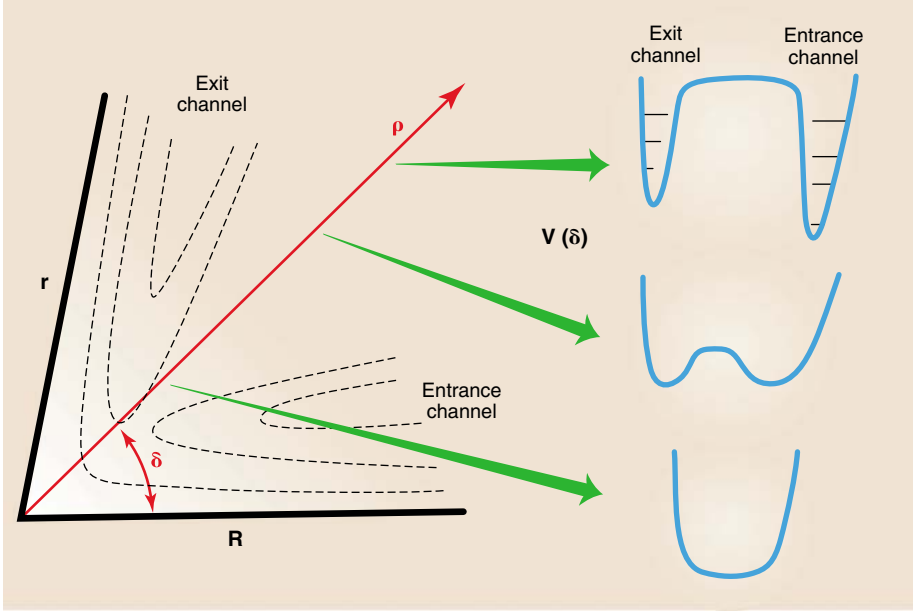


Fig. 1. Schematic diagram of a potential energy surface for a colinear atom + diatom reaction, with bonds being broken (*r*) and formed (*R*). A transformation is made of (*R*, *r*) to the polar (hyperspherical) coordinates (ρ, δ). The potentials *V*(δ) in the angular variable δ for three different values of ρ are also shown. The quantum wave function is integrated along the hyperradius ρ and, for fixed values of ρ, vibrational energy levels are obtained for each value of *V*(δ). When ρ is large, these levels correspond to the vibrational states of reactants or products.

Department of Physical and Theoretical Chemistry, University of Oxford, Oxford OX1 3QZ, UK. E-mail: david.clary@chem.ox.ac.uk

provide a multidimensional mathematical function for the potential energy surface. The final task, then, is to use these potential surfaces in the solution of the Schrödinger equation for the nuclei. These procedures provide all the fundamental information needed to predict results for comparison with essentially any experiment on chemical reactions.

Three-Atom Reactions

The simplest reactions are those involving three atoms

$$A + BC \rightarrow AB + C \qquad (1)$$

On these systems, it is now possible to perform full quantum scattering calculations with the use of a general computer code called ABC (*9*). The method used in this code involves hyperspherical coordinates (*10*, *11*), which reduce to polar coordinates in two dimensions and allow for the vibrational states of reactants and products to be defined naturally as part of the calculational procedure (Fig. 1). This theory can be extended to full dimensions by including functions that describe molecular rotations (*12*).

This theory has been applied to a range of three-atom systems. The reactions between H_2 and F or H atoms provide the most detailed comparisons between theory and experiment as the potential energy surfaces for these reactions, which have a relatively small number of electrons, are readily calculated accurately (*1*, *2*, *13*–*15*). The agreement between theory and experiment for these reactions is often exceptional, and this agreement is expected for what is essentially a numerically exact solution of the Schrödinger equation.

This theory has been extended to compute reactive transitions from different electronic states of the F atom (*16*). High-resolution molecular-beam experiments on

$$F + D_2 \rightarrow DF(v) + D \qquad (2)$$

measure the differential cross sections—that is, the flux of reaction products going into different scattering angles and product DF in vibrational state v for the initial levels $F(^2P_{1/2})$ and $F(^2P_{3/2})$. The computational method includes the spin-orbit coupling between the two electronic states. As shown in Fig. 2, the agreement between theory and experiment is remarkable (*16*). The large peaks seen in the differential cross sections are associated with scattering resonances that have been predicted by theory for many years for this reaction (*17*) but only recently confirmed in these and related experiments (*15*, *16*). The resonances can occur when the topology of the potential surface and couplings between the vibrational states allow for temporary vibrational excitation of the FD_2 complex formed in the reaction. An unexpected finding in this study is that $F(^2P_{1/2})$ is more reactive than $F(^2P_{3/2})$ at low collision energies, although the reaction from $F(^2P_{1/2})$ is forbidden within the Born-Oppenheimer approximation (*16*). This outcome demonstrates the importance of including spin-orbit terms that couple together the $F(^2P_{1/2})$ and $F(^2P_{3/2})$ states in the theoretical treatment of a reaction of this type.

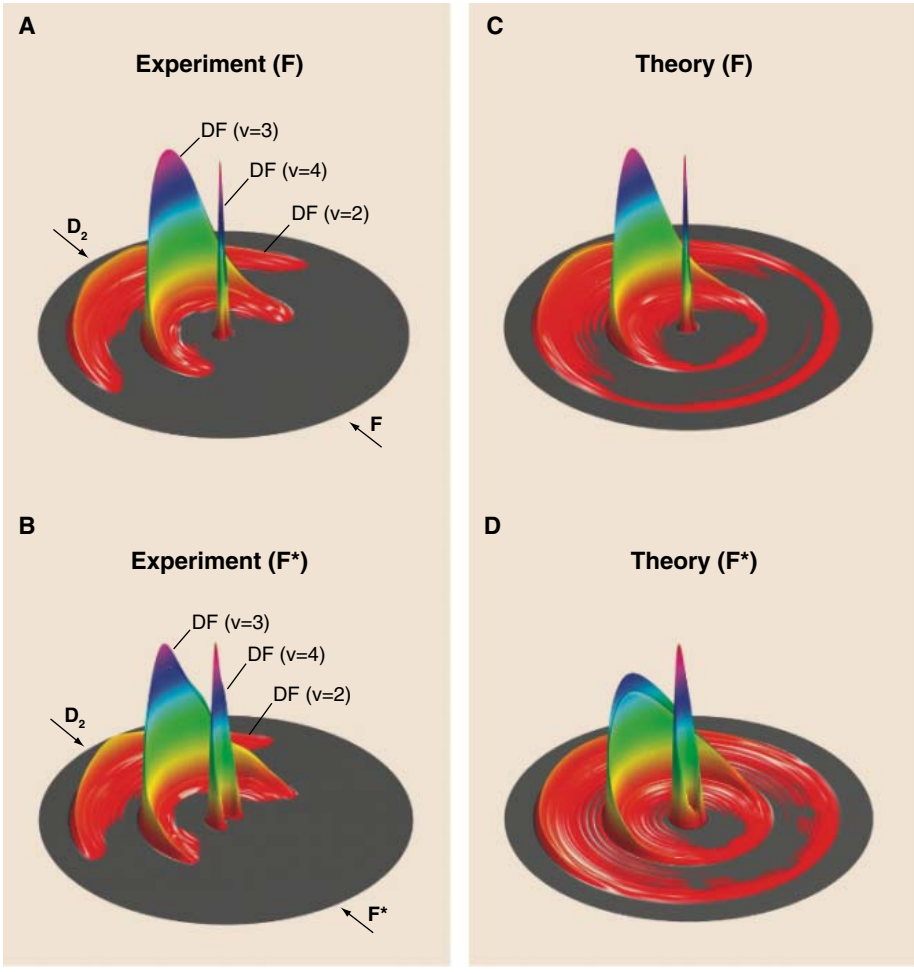


Fig. 2. Excellent agreement is found between the experimental (**A** and **B**) and theoretical (**C** and **D**) D atom product flux for the $F(^2P_{3/2})/F(^2P_{3/2}) + D_2$ ($j = 0$) reaction at a collision energy of 0.48 kcal/mol. The concentric circles correspond to different product ro-vibrational states (j, v) of the DF product. The direction of the F atom beam defines the forward scattering direction. The asterisk indicates the electronically excited $F(^2P_{3/2})$ state. Figure reproduced from (*16*).

Reactions with several maxima and minima in the potential energy surface, such as those involving ions or free radicals, can cause difficulties when performing electronic structure calculations and fitting the required potential energy surfaces. For these reactions, "direct dynamics" methods are sometimes used. These methods involve calculating electronic energies directly from an ab initio code whenever they are required (*18*). These types of calculations are often done in molecular dynamics computations or transition-state theory applications to chemical reactions. However, the direct dynamics method is very difficult to apply in quantum dynamics calculations, due to the large numbers of electronic energies that are needed because the quantum wave functions extend over very large regions of configuration space.

In addition, the quantum dynamics calculations can be particularly challenging for reactions with deep minima in the potential surface that require the coupling of very large numbers of quantum states that become energetically accessible during the reaction. For example, the three-atom reaction

$$O(^1D) + H_2 \rightarrow OH + H \qquad (3)$$

is very exothermic, and part of the potential energy surface includes the electronic ground state of the H_2O molecule (*19*). In addition, the minimum energy reaction path has a geometry in which the $O(^1D)$ atom inserts itself between the two hydrogen atoms of H_2. Most quantum dynamics calculations on this reaction have been

performed with an alternative method that involves solving the time-dependent Schrödinger equation (*20*–*22*). These "wave packet" calculations can be done separately for each initial state of the reactant H_2 molecule, and this can reduce the computational expense. Furthermore, images of reactive flux developing as a function of time provide new information on the detailed reaction mechanism. The method also has the advantage that it provides information for a large number of collision energies from a single dynamical calculation and can be extended in a fairly straightforward way to excited electronic states, as is required in the most detailed calculations on the $O(^1D) + H_2$ reaction (*21*). These calculations, along with the detailed comparison with the results of crossed molecular-beam experiments, have established a mechanism of the reaction that involves an insertion into the middle of the H_2 bond via a long-lived complex at low collision energies. In contrast, at higher energies the reaction involves a more direct abstraction of an H atom (*3*, *23*).

Reactions Involving Four or More Atoms

One of the major challenges is to extend the success of the theory obtained for atom + diatom reactions to systems involving more than three atoms. The benchmark four-atom reaction is

$$OH + H_2 \rightarrow H_2O + H \qquad (4)$$

The hyperspherical coordinate method has been extended to this case (*24*), and predictions (*25*) of vibrational mode selectivity effects in the HOD product of the $OH + D_2$ reaction were confirmed in subsequent experiments (*26*). These calculations, however, are difficult to perform because of the very large number of quantum states that need to be coupled together in this reaction. The more extensive calculations on this reaction have been carried out with the use of the wave packet method (*27*, *28*). However, even for a six-atom reaction such as

$$H + CH_4 \rightarrow CH_3 + H \qquad (5)$$

converged wave packet calculations in full dimensions are very hard to carry out and are normally done more approximately by reducing the number of degrees of freedom coupled together in the calculation (*29*). In addition, obtaining a full-dimensional potential energy surface for a reaction such as this to use in quantum dynamics calculations is a complicated process, although there has been promising recent progress that exploits the permutations of identical atoms (*30*).

The fitting of potential energy surfaces can be avoided completely if all electronic energies can be calculated from first principles whenever they are needed. A method has been developed that allows for the construction of effective potential surfaces in reduced dimensionality through a small number of accurate electronic structure calculations (*31*). These surfaces are then used in time-independent quantum scattering calculations in hyperspherical coordinates, treating only the bonds being broken and formed in a reaction explicitly. This procedure has been carried out for various reactions involving polyatomic molecules; a good example is the reaction of H atoms with propane ($H + C_3H_8$), which yields H_2 and two product radicals, isopropyl (CH_3CHCH_3) and *n*-propyl ($CH_2CH_2CH_3$). Here, quantum tunneling is found to be significant, and the calculated branching ratio for producing the two chemically distinct product radicals in the reaction agrees well with the experimental results (*32*).

Another major area of research in chemical reaction theory involves exploiting computational techniques that allow for the coupling together of large numbers of degrees of freedom. This includes coupled coherent states (*33*) and the multiconfigurational time-dependent Hartree (MCTDH) approach (*34*). MCTDH writes the quantum wave function as a product of single-particle and time-dependent functions. The equations of motion are a set of complicated nonlinear differential equations, but efficient integration schemes have been developed to solve them (*35*).

One of the most impressive applications of MCTDH is its combination with a method for calculating the quantum fluxes in a chemical reaction. This procedure allows for reactive rate constants to be calculated very efficiently for multidimensional systems. For the challenging $H + CH_4$ reaction and its reverse ($H_2 + CH_3$), converged rate constants have been calculated in full dimensions (*36*). By exploiting new methods for fitting potential surfaces (*37*), an optimum and general technique for calculating rate constants has been developed that has considerable promise for more complicated reactions. However, there have been very few applications as yet of the MCTDH method to state-to-state reaction dynamics.

Reactions of Biological Importance

Going beyond reactions of small molecules in the gas phase, quantum effects such as tunneling are known to be critical in reactions of biological importance. To calculate rate coefficients for systems of this kind, the most common procedure is to apply a variational transition-state approximation with a tunneling correction; this has been useful in explaining some subtle isotope effects in hydrogen-transfer reactions catalyzed by enzymes (*38*). Other approximations that use hybrid quantum and classical dynamics methods have promise for these systems (*39*), as do methods that apply semiclassical approaches (*40*). A major challenge in the next decade will be to extend the quantum dynamical theories to these types of reactions to yield more accurate results.

References and Notes

1. W. Hu, G. C. Schatz, *J. Chem. Phys.* **125**, 132301 (2006).
2. S. C. Althorpe, D. C. Clary, *Annu. Rev. Phys. Chem.* **54**, 493 (2003).
3. X. Yang, *Annu. Rev. Phys. Chem.* **58**, 433 (2007).
4. D. C. Clary, *Annu. Rev. Phys. Chem.* **41**, 61 (1990).
5. J. Troe, *J. Chem. Soc. Faraday Trans.* **90**, 2303 (1994).
6. T. Schlik, *Molecular Modeling and Simulation* (Springer, New York, 2002).
7. A. I. Krylov, *Annu. Rev. Phys. Chem.* **59**, 433 (2008).
8. R. A. Friesner, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6648 (2005).
9. D. Skouteris, J. F. Castillo, D. E. Manolopoulos, *Comput. Phys. Commun.* **133**, 128 (2000).
10. A. Kuppermann, J. A. Kaye, J. P. Dwyer, *Chem. Phys. Lett.* **74**, 257 (1980).
11. J. Romelt, *Chem. Phys. Lett.* **74**, 263 (1980).
12. R. T. Pack, G. A. Parker, *J. Chem. Phys.* **87**, 3888 (1987).
13. D. Dai *et al.*, *Science* **300**, 1730 (2003).
14. R. T. Skodje, X. M. Yang, *Int. Rev. Phys. Chem.* **23**, 253 (2004).
15. X. Wang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6227 (2008).
16. M. Qiu *et al.*, *Science* **311**, 1440 (2006).
17. G. C. Schatz, J. M. Bowman, A. Kuppermann, *J. Chem. Phys.* **58**, 4023 (1973).
18. J. G. Lopez *et al.*, *J. Am. Chem. Soc.* **129**, 9976 (2007).
19. R. Schinke, W. A. Lester, *J. Chem. Phys.* **72**, 3754 (1980).
20. X. Liu, C. C. Wang, S. A. Harich, X. Yang, *Phys. Rev. Lett.* **89**, 133201 (2002).
21. S. K. Gray, C. Petrongolo, K. Drukker, G. C. Schatz, *J. Phys. Chem. A* **103**, 9448 (1999).
22. K. Yuan *et al.*, *Phys. Rev. Lett.* **96**, 103202 (2006).
23. F. J. Aoiz *et al.*, *J. Chem. Phys.* **116**, 10692 (2002).
24. D. C. Clary, *J. Chem. Phys.* **95**, 7298 (1991).
25. S. K. Pogrebnya, J. Palma, D. C. Clary, J. Echave, *Phys. Chem. Chem. Phys.* **2**, 693 (2000).
26. B. R. Strazisar, C. Lin, H. F. Davis, *Science* **290**, 958 (2000).
27. D. H. Zhang, J. Z. H. Zhang, *J. Chem. Phys.* **101**, 1146 (1994).
28. D. H. Zhang, *J. Chem. Phys.* **125**, 133102 (2006).
29. M. H. Yang, D. H. Zhang, S. Y. Lee, *J. Chem. Phys.* **117**, 9539 (2002).
30. X. Zhang, B. Braams, J. M. Bowman, *J. Chem. Phys.* **124**, 021104 (2006).
31. B. N. Kerkeni, D. C. Clary, *Phys. Chem. Chem. Phys.* **8**, 917 (2006).
32. B. N. Kerkeni, D. C. Clary, *Mol. Phys.* **103**, 1745 (2005).
33. D. V. Shalashilin, M. S. Child, *J. Chem. Phys.* **115**, 5367 (2001).
34. H.-D. Meyer, U. Manthe, L. S. Cederbaum, *Chem. Phys. Lett.* **165**, 73 (1990).
35. H.-D. Meyer, G. A. Worth, *Theor. Chem. Acc.* **109**, 251 (2003).
36. T. Wu, H.-J. Werner, U. Manthe, *Science* **306**, 2227 (2004).
37. M. A. Collins, *Theor. Chem. Acc.* **108**, 313 (2002).
38. A. Dybala-Defratyka, P. Paneth, R. Banerjee, D. G. Truhlar, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 10774 (2007).
39. Q. Wang, S. Hammes-Schiffer, *J. Chem. Phys.* **125**, 184102 (2006).
40. W. H. Miller, *J. Chem. Phys.* **125**, 132305 (2006).

41. This work was supported by the Engineering and Physical Sciences Research Council and the Office of Naval Research under grant number N00014-05-1-0460.

10.1126/science.1157718

PERSPECTIVE

Insights into Current Limitations of Density Functional Theory

Aron J. Cohen, Paula Mori-Sánchez, Weitao Yang*

Density functional theory of electronic structure is widely and successfully applied in simulations throughout engineering and sciences. However, for many predicted properties, there are spectacular failures that can be traced to the delocalization error and static correlation error of commonly used approximations. These errors can be characterized and understood through the perspective of fractional charges and fractional spins introduced recently. Reducing these errors will open new frontiers for applications of density functional theory.

Interactions between electrons determine the structure and properties of matter from molecules to solids. To describe interacting electrons, the extremely simple three-dimensional electron density can be used as the basic variable within density functional theory (DFT) (*1*, *2*), negating the need in many cases for the massively complex many-dimensional wave function.

Kohn noted in his Nobel lecture that DFT "has been most useful for systems of very many electrons where wave function methods encounter and are stopped by the exponential wall" (*3*). The beauty of DFT is that its formalism is exact yet efficient, with one determinant describing the electron density—all of the complexity is hidden in one term, the exchange-correlation functional. This term holds the key to the success or failure of DFT. Exchange arises from antisymmetry due to the Pauli exclusion principle, and correlation accounts for the remaining complicated many-body effects that need many determinants to be fully described. However, the form of exchange-correlation in terms of the density remains unknown and it is necessary to use approximations, so that in many cases DFT contains semi-empirical parameters. The great success of DFT is that simple approximations perform remarkably well for a wide range of problems in chemistry and physics (*4*–*6*), particularly for prediction of the structure and thermodynamic properties of molecules and solids.

Despite the widespread popularity and success of DFT, its application can still suffer from large pervasive errors that cause qualitative failures in predicted properties. These failures are not breakdowns of the theory itself but are only due to deficiencies of the currently used approximate exchange-correlation functionals. A systematic approach for constructing functionals that are universally applicable is a hard problem and has remained elusive.

A possible path forward is to have a deeper look at the errors embedded in currently used functionals to determine the origin of their pathologies at the most basic level. Recent work has traced many of the errors in calculations to violations of conditions of the exact functional (*7*, *8*). These violations present themselves in extremely simple model atoms, which can be used for diagnosis, and more importantly, in many interesting and complex chemical and physical systems. Identifying and understanding the basic errors offer a much needed path for the development of functionals, as well as a useful insight into potential pitfalls for practical applications.

What are some of the major failures in DFT calculations? First, they underestimate the barriers of chemical reactions, the band gaps of materials, the energies of dissociating molecular ions, and charge transfer excitation energies. They also overestimate the binding energies of charge transfer complexes and the response to an electric field in molecules and materials. Surprisingly, all of these diverse issues share the same root—the delocalization error of approximate functionals, due to the dominating Coulomb term that pushes electrons apart. This error can be understood from a perspective that invokes fractional charges (*7*, *9*). Furthermore, typical DFT calculations fail to describe degenerate or near-degenerate states, such as arise in transition metal systems, the breaking of chemical bonds, and strongly correlated materials. All of these problems are merely manifestations of another common error—the static correlation error of approximate functionals. This problem arises because of the difficulty in using the electron density to describe the interaction of degen-

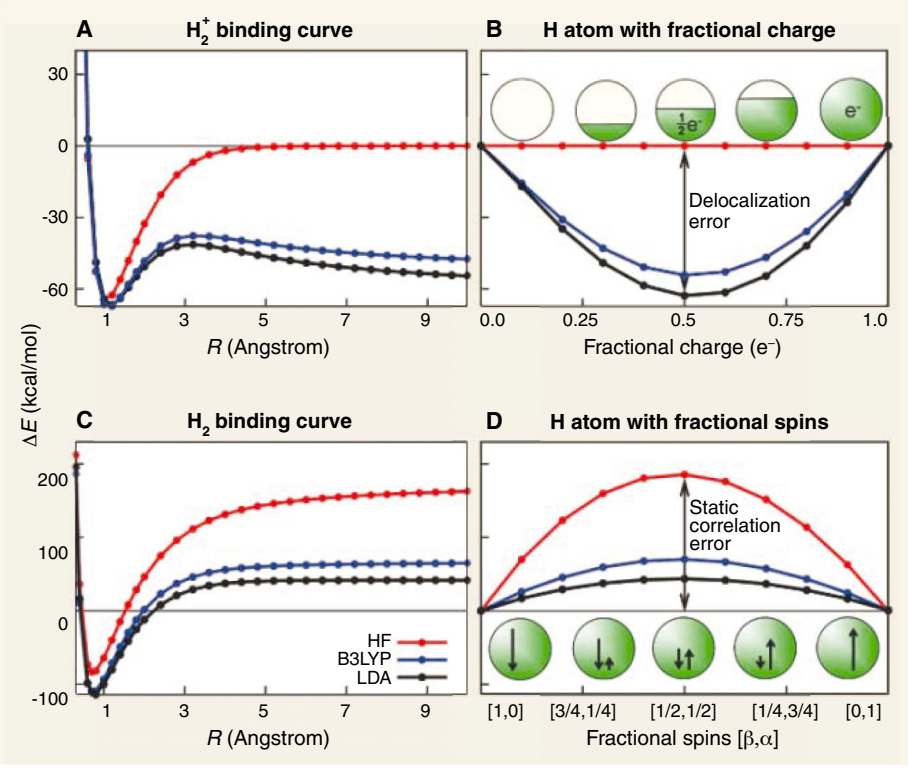


Fig. 1. DFT approximations fail. The dissociation of H_2^+ molecule (**A**) and H_2 molecule (**C**) are shown for calculations with approximate functionals: Hartree-Fock (HF), local density approximation (LDA), and B3LYP. Although DFT gives good bonding structures, errors increase with the bond length. The huge errors at dissociation of H_2^+ exactly match the error of atoms with fractional charges (**B**), and for H_2 they exactly match the error of atoms with fractional spins (**D**). The understanding of these failures leads to the characterization of the delocalization error and static correlation error that are pervasive throughout chemistry and physics, explaining a host of problems with currently used exchange-correlation functionals.

Department of Chemistry, Duke University, Durham, NC 27708, USA.

*To whom correspondence should be addressed. E-mail: weitao.yang@duke.edu

erate states and can be understood in another perspective, that of fractional spins (*8*).

Delocalization Error and Fractional Charges

To understand the delocalization error inherent in approximate functionals and its effects, we can consider the simplest possible molecule, H_2^+. Functionals describe the chemical bond well, but fail dramatically as the molecule is stretched (Fig. 1A). At the dissociation limit, the result is two H atoms with half an electron each, and an energy that is much too low. The H atom with half an electron is a curious object; what do we know, in general, about atoms with fractional numbers of electrons? The exact energy of the atom as a function of the charge is a straight-line interpolation between the integers (*10*), because of the discrete nature of electrons. However, approximate functionals are incorrectly convex between the integers (Fig. 2A), with large errors, and they predict an energy that is much too low for fractional charges (Fig. 1B).

Stretched H_2^+ clearly shows that the energy is too low if the electron is delocalized over the two centers. This reflects the delocalization error, which refers to the tendency of approximate functionals to spread out the electron density artificially. This phenomenon is related to a familiar term in the literature, the "self-interaction error" (*11*, *12*), well defined only for one-electron systems, with analogy made for many-electron systems (*13*, *14*). However, the erroneous convex behavior of the energy as a function of the fractional charge is best characterized as the delocalization error, as it captures the physical nature of the problem (*7*).

What are the consequences of the delocalization error? In chemical reactions, approximate functionals predict transition state energies and hence reaction barriers that are much too low (*15*), because the transition state resembles that of stretched H_2^+ and has electrons delocalized over more than one center (*12*). In an applied electric field, approximate functionals give the system an energy that is too low and a polarizability that is too high because fractional charges appear at the edges of the molecule (*16*). Approximate DFT can also predict unphysical charge transfer between the molecule and the metal leads in molecular electronic devices, and cause an overestimation of electronic conductance (*17*, *18*). When adding or removing an electron from a system, approximate functionals produce an overly disperse distribution for the added electron or hole (Fig. 2B). This effect can be magnified in large systems because the delocalization error increases with system size. Thus, the delocalization error may be a particular concern for cluster calculations that attempt to approach the bulk limit.

For the exact functional, its straight-line behavior (Fig. 2A) means that the derivatives of the energy with respect to fractional charge give the ionization energy and the electron affinity (*9*). Functionals with delocalization error get these derivatives wrong because of incorrect convex behavior. This explains the well-known underestimation of the band gap. Functionals with incorrect concave behavior, like the Hartree-Fock approximation (which is an orbital functional within DFT), have the opposite characteristic and suffer from a localization error that saturates with system size (*7*). Hybrid functionals contain both convex and concave components and may benefit from error cancellation in some cases, thus reproducing good band gaps for solids with midsize gaps; however, they do not provide general solutions.

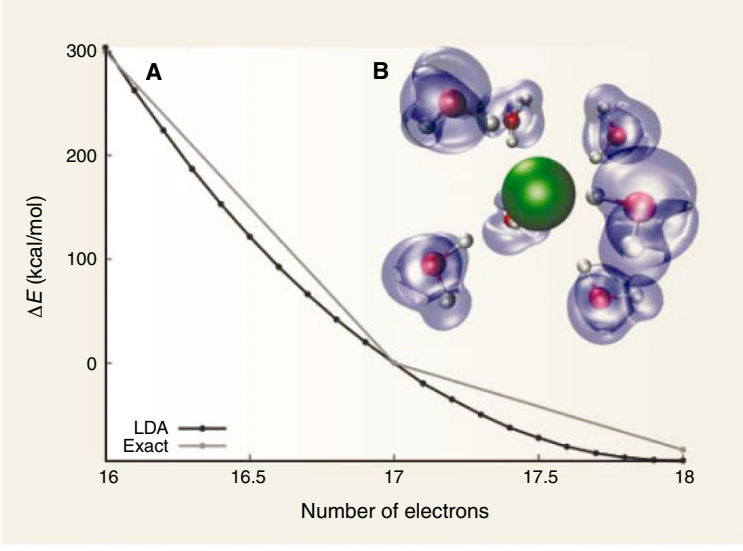


Fig. 2. Seeing the delocalization error. The incorrect convex behavior of the LDA energy is seen for the Cl atom with fractional numbers of electrons (**A**). Despite good agreement with the exact values at the integer points, the convex behavior means that a lower energy can be achieved with fractional charges. This leads to an unphysical delocalization of the electrons in a simulation of the solvated anion, as shown (in blue) by the difference of the LDA density between the solvated Cl and solvated Cl^- (**B**).

Static Correlation Error and Fractional Spins

Next, we consider the apparently simple problem of stretching H_2. Approximate functionals describe the covalent bond well, but the energy is severely overestimated upon dissociation (Fig. 1C). This well-known problem is attributed to static correlation that arises in situations with degeneracy or near-degeneracy (*19*–*21*), as in transition metal chemistry and strongly correlated systems. These are very complicated problems with no clear solution within DFT. Nonetheless, recent studies offer insight and promise (*8*).

To analyze the error, consider one half of the closed-shell H_2 molecule at the dissociation limit: an H atom with half a spin-up electron and half a spin-down electron, which is an exotic system with fractional spins. The exact energy of this system should be the same as that of the normal H atom in an integer-spin state. More generally, the constancy condition for the exact functional (*8*) requires that systems with fractional spins have an energy equal to that of the normal-spin states. However, approximate functionals do not accurately describe the interaction between the degenerate spin-up and spin-down states and have a massive error for these fractional-spin states. In the case of the H atom, the overestimation of this energy exactly matches the error of stretched H_2 (Fig. 1D). The violation of the constancy condition for fractional spins gives rise to the static correlation error and explains the difficulty in using the electron density to describe degenerate states.

Stretched H_2 and similar problems are dealt with in traditional approaches using multideterminant wave functions. For larger systems with many degeneracies, the number of determinants quickly becomes unmanageable. In contrast, DFT is based on a single determinant, and the above solution may not apply. However, the concept of fractional spins formulates the very same problem in a different light that illuminates the error of approximate functionals and, moreover, indicates that all that is needed in DFT to address the problem is to impose the constancy condition on an approximate functional.

Static correlation is pervasive. Breaking any chemical bond leads to failures similar to those seen in H_2, with a static correlation error that is proportional to the number of electron pairs involved. For transition metal dimers, which have multiple bonds and hence multiple degeneracies, the static correlation error dominates. As such, DFT calculations cannot accurately describe the chemical bonding in metal dimers. Another simple example is the calculation of the open-shell boron atom, which has a threefold spatial degeneracy among the p orbitals. DFT calculations of the spherical density give an excessively high energy because of multiple fractional-spin occupations.

Magnetic properties, such as the magnetic coupling constants of cuprates, rely on the accurate calculation of both singlet and triplet states.

At present, the accurate calculation of closed-shell singlets is not always possible in DFT because of static correlation error, which has led to the routine use of broken-symmetry open-shell singlets. The use of broken-symmetry solutions in all of the above cases can be understood as only attempting to circumvent the static correlation error of the functional by using a different state that avoids fractional spins. Static correlation error is particularly important in strongly correlated systems, as exemplified by Mott insulators, including transition metal oxides, where DFT is unable to predict the correct insulating state. Related problems exist in high-temperature superconductors. Functionals that include Hartree-Fock, such as hybrid functionals, have a larger error than normal functionals (Fig. 1D) and are expected to perform even worse on static correlation problems.

DFT can suffer from large errors that manifest themselves in a wide range of applications. However, they also appear in the simplest systems, even atoms, which can be used to understand and diagnose problems. In this manner, perspectives based on the behavior of the energy for fractional charges and fractional spins give insight into the errors of functionals and illuminate a path for further development. Indeed, new functionals have been constructed to minimize the delocalization error, and these are found to improve many related predictions (*22*), but there is still much work to do. It is important to describe errors of functionals in terms of violations of requirements of the exact functional; for example, the delocalization error originates from the violation of linearity of the energy as a function of fractional charges, and the static correlation error emerges from the violation of constancy of the energy as a function of fractional spins. A clear understanding of the errors from the most basic principles will enable the development of functionals to open new frontiers for applications of DFT.

References and Notes

1. W. Kohn, L. Sham, *Phys. Rev. A* **140**, 1133 (1965).
2. R. G. Parr, W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford Univ. Press, New York, 1989).
3. W. Kohn, *Rev. Mod. Phys.* **71**, 1253 (1999).
4. A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993).
5. C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).
6. J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
7. P. Mori-Sánchez, A. J. Cohen, W. Yang, *Phys. Rev. Lett.* **100**, 146401 (2008).
8. A. J. Cohen, P. Mori-Sánchez, W. Yang, http://arXiv.org/abs/0805.1724.
9. A. J. Cohen, P. Mori-Sánchez, W. Yang, *Phys. Rev. B* **77**, 115123 (2008).
10. J. P. Perdew, R. G. Parr, M. Levy, J. L. Balduz Jr., *Phys. Rev. Lett.* **49**, 1691 (1982).
11. J. P. Perdew, A. Zunger, *Phys. Rev. B* **23**, 5048 (1981).
12. Y. Zhang, W. Yang, *J. Chem. Phys.* **109**, 2604 (1998).
13. P. Mori-Sánchez, A. J. Cohen, W. Yang, *J. Chem. Phys.* **125**, 201102 (2006).
14. A. Ruzsinszky, J. P. Perdew, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, *J. Chem. Phys.* **126**, 104102 (2007).
15. Y. Zhao, B. J. Lynch, D. G. Truhlar, *J. Phys. Chem. A* **108**, 2715 (2004).
16. P. Mori-Sanchez, Q. Wu, W. Yang, *J. Chem. Phys.* **119**, 11001 (2003).
17. C. Toher, A. Filippetti, S. Sanvito, K. Burke, *Phys. Rev. Lett.* **95**, 146402 (2005).
18. S.-H. Ke, H. U. Baranger, W. Yang, *J. Chem. Phys.* **126**, 201102 (2007).
19. A. Savin, in *Recent Developments and Applications of Modern Density Functional Theory*, J. M. Seminario, Ed. (Elsevier, Amsterdam, 1996), p. 327.
20. E. J. Baerends, *Phys. Rev. Lett.* **87**, 133004 (2001).
21. A. D. Becke, *J. Chem. Phys.* **119**, 2972 (2003).
22. A. J. Cohen, P. Mori-Sánchez, W. Yang, *J. Chem. Phys.* **126**, 191109 (2007).

23. We acknowledge the NSF for financial support and D. N. Beratan and J. M. Parks for help in preparation of the manuscript.

10.1126/science.1158722

PERSPECTIVE

Frontiers in Surface Scattering Simulations

Geert-Jan Kroes

Theorists have recently made substantial progress in simulating reactive molecule-metal surface scattering but still face major challenges. The grand challenge is to develop an approach that enables accurate predictive calculations of reactions involving electronically excited states with potential curve crossings. This challenge is all the more daunting because collisions involving molecules heavier than H_2 may be accompanied by substantial energy exchange with the surface vibrations (phonons), and because an electronic structure approach that allows molecule-surface interaction energies to be computed with chemical accuracy (1 kilocalorie per mole) is not yet available even for the electronic ground state of molecule-metal surface systems.

The goal of molecule-surface reaction dynamicists is to determine in detail how molecules react on surfaces. On surfaces, a wealth of physical and chemical processes are possible (Fig. 1): Examples include diffractive scattering of light atoms and molecules (*1*), in which the projectile's translational momentum parallel to the surface is altered in a quantized way; vibrationally inelastic scattering (*2*, *3*), in which the molecule either accumulates or dissipates springlike internal vibrational energy from the collision; and dissociative chemisorption (*1*, *4*–*6*), in which a bond in the incident molecule breaks and the fragments bond chemically to the surface. Theorists study these processes to understand how the molecule-surface interaction affects the elementary reactions that together constitute heterogeneously catalyzed processes. Such reactions are important; the production of most synthetic chemical compounds involves heterogeneous catalysis. One of the achievements recognized by the 2007 Nobel Prize in Chemistry, awarded to G. Ertl, was the unraveling of the fundamental mechanism of the Haber-Bosch process, by which vast quantities of ammonia, a raw material for fertilizer, are produced.

Selecting a model for a surface reaction requires decisions about how to treat the electronic and nuclear degrees of freedom. In the Born-Oppenheimer (BO) approximation, the assumption is made that the electrons adapt instantaneously to the motion of the nuclei, so that the molecule-surface interaction can be represented using just one potential energy surface (PES). This assumption does not hold true for nonadiabatic processes involving, for instance, spin changes (*6*) or electron transfer (*7*). The modeling of such processes requires computing a diabatic PES (a diabatic state is characterized by a physical feature other than electronic energy—for instance, the molecular spin state) or multiple PESs with couplings connecting them. Low-energy electron-hole pair excitations, which form another example of a nonadiabatic process, can occur in scattering from a metal in which the highest occupied electronic band is only partially filled; these can be treated by replacing them with effective frictions [molecular dynamics with electronic friction (MDEF)] (*8*). The equations of motion of the nuclei can be solved using the time-independent or time-dependent Schrödinger equation to treat all molecular degrees of freedom [often the approach taken for dissociative chemisorption of a light molecule such as H_2 (*9*)] or to treat only the most important molecular degrees of freedom [often applied to reactions of heavy diatomic (*7*) or polyatomic (*10*) molecules]. Surface vibrations, or phonons, can be incorporated

Leiden Institute of Chemistry, Leiden University, Gorlaeus Laboratories, Post Office Box 9502, 2300 RA Leiden, Netherlands.

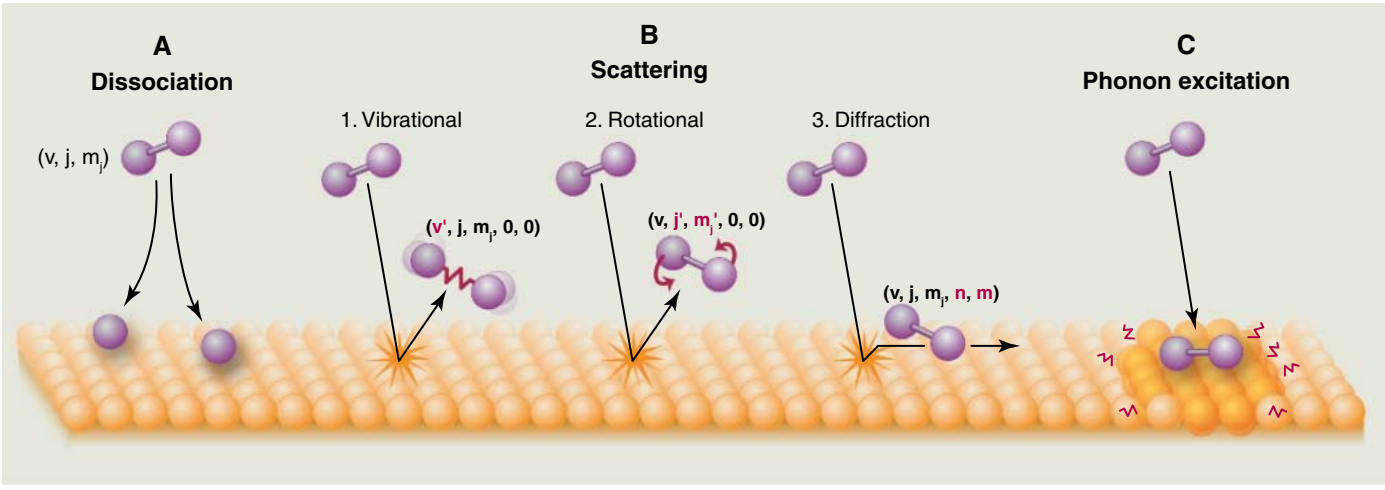


Fig. 1. Schematic representations of some of the processes that can be probed in the scattering of a diatomic molecule from a reactive surface. A molecule approaching the surface in a specific initial internal state characterized by the vibrational quantum number *v*, the rotational angular momentum quantum number *j*, and the magnetic rotational quantum number m_j can either react (**A**) or scatter back to the gas phase (**B**) in a vibrationally excited state (1) or a rotationally excited state (2) or with a quantized change in the momentum parallel to the surface [diffraction (3)]. (**C**) All of these processes can occur in coincidence with excitation of the phonons and with electron-hole pair excitations (not shown).

classically using ab initio molecular dynamics (*4*), or a single phonon can be treated quantum-mechanically with a reduced dimensionality treatment of the molecular degrees of freedom (*11*).

Recent experiments have provided evidence that reactive (*12*, *13*) and nonreactive (*13*, *14*) molecule-metal surface scattering can be accompanied by substantial nonadiabatic effects. Therefore, much recent research has focused on whether reactions can be found that can accurately be described by the BO approximation. How can one test whether nonadiabatic events such as electron-hole pair excitation, which can act as a friction phenomenon in scattering (*8*), affect a simple reaction such as the dissociation of H_2 on a metal surface? One way is to check whether a treatment employing a single PES produces accurate results for both reaction (which friction can affect on the way to the barrier) and scattering (which friction can also affect on the way back). Calculations on H_2 + Pt(111) performed within the BO approximation (*1*) gave accurate results for both reaction (Fig. 2A) and diffractive scattering (fig.S1), making it very unlikely that electron-hole pairs substantially affect reaction in this system. The BO approximation accurately describes the reaction of H_2 on metal surfaces because the reaction path does not involve a deep molecular chemisorption well in front of the barrier and because H_2 has a low affinity for electrons (*1*). This makes H_2 interacting with metal surfaces test systems for developing accurate electronic structure methods for computing molecule-surface interaction energies, and this is important because although the high-level ab initio methods accessible to gas-phase theorists provide chemical accuracy (1 kcal/mol ≈ 4.2 kJ/mol) of reaction barrier heights, errors associated with molecule-metal surface reactions are significantly higher. A recent benchmark study addressing 114 gas-phase reaction barrier heights yielded a mean unsigned error of 5 kcal/mol for the presently used generalized gradient approximation (GGA) functionals of density functional theory [DFT, the method of choice for molecule-metal surface interaction energies (*15*)], which can be brought down by about 1 kcal/mol by changing to a so-called meta-GGA functional (*16*). Although this accuracy represents an enormous improvement over what was possible 20 years ago, it still fails to offer adequate predictions for reaction probabilities that can be measured in molecular beam experiments. Therefore, one major challenge facing theorists is to come up with more accurate computational schemes to compute interaction energies for molecule-metal surface systems in their electronic ground state. Possible ways of achieving improvements include embedding schemes in which the molecule and part of the metal surface are treated by a high-level ab initio method while DFT is used to model the density of the surrounding metal environment (*17*).

In the ongoing debate on adiabaticity, energy dissipation to electron-hole pairs was proposed to dominate the reaction of N_2 on Ru(0001) (*18*). This proposition was motivated by the following findings: (i) the dissociative chemisorption probability is only 1% at collision energies (≈400 kJ/mol), twice as high as the reaction barrier (≈200 kJ/mol), and (ii) low-dimensional (2 + 1D) quasi-classical trajectory calculations that modeled motion in the N-N and molecule-surface distance coordinates produced reaction probabilities too high by two orders of magnitude (Fig. 2B) (*19*). However, quasi-classical trajectory calculations (*5*) modeling all six N_2 molecular degrees of freedom and performed within the BO approximation gave better agreement with experimental data than low-dimensional (2 + 2D) trajectory calculations that also considered electron-hole pair excitation (*18*) (Fig. 2B), showing that, for N_2-metal systems, modeling motion in all molecular degrees of freedom can be more relevant to achieving accuracy than modeling electron-hole pair excitation.

To complicate the treatment of molecule-surface reactions even further, reactive scattering may be affected by energy exchange with the phonons (*10*). The effect of phonons on rotationally inelastic scattering of D_2 from Cu(111) (*11*) and H_2 from Pd(111) (*19*) has been investigated with quantum dynamics (*11*) and classical trajectory (*19*) calculations. Both sets of calculations show a surface-temperature dependence having an Arrhenius form with the activation energy depending on the collision energy, in good agreement with molecular beam experimental data (fig. S2) (*20*). This agreement could be achieved while using an approximate model for surface motion [the surface oscillator (SO) model]. In this model, the surface atom with which the molecule interacts vibrates with a frequency (or frequencies) appropriate for surface phonons, and the effect of this motion on the molecule-surface interaction is described by replacing $\boldsymbol{R}_p$ with $\boldsymbol{R}_p - \boldsymbol{R}_t$ in the expression for the PES, where $\boldsymbol{R}_p$ and $\boldsymbol{R}_t$ are the vectors defining the position of the projectile molecule and the target surface atom, respectively.

Recent quantum-dynamical calculations on the dissociative chemisorption of CH_4 on Ni(111) (*10*) address the validity of the second approximation of the SO model (simply replacing $\boldsymbol{R}_p$ with $\boldsymbol{R}_p - \boldsymbol{R}_t$ in the expression for the PES) to describe the effect of phonons on reaction. The calculations showed that the reaction

barrier can be lowered by a surface reconstruction, in which the surface Ni atom nearest to CH_4 moves out of the surface. The lattice-reconstruction model in (*10*), in which the dependence of the molecule-surface interaction on the motion of the surface atom was explicitly computed and modeled, yields reaction probabilities that, in agreement with experimental data, increase with surface temperature and are much higher than those obtained for the simple SO model (Fig. 2C). The reaction probabilities were also much higher than those obtained for a static surface (SS) model (Fig. 2C). This emphasizes the need to correctly model surface vibrations to obtain accurate results for heavy molecules such as CH_4, the mass of which is much more similar to that of a typical surface atom than that of the light H_2 molecule. In fact, accurately treating phonons when dealing with molecule-surface reactions represents another major challenge. Research has shown that, although the common-practice treatment of phonons through the oscillation of a single surface atom (the first approximation made in the SO model) may work well in certain instances (*11*, *19*), this practice is actually somewhat suspect (*21*). Ab initio molecular dynamics treatments that also simulate surface vibrations (*4*), which circumvent this approximation, face issues concerning the applicability of classical mechanics to vibrational energy transfer.

A second challenge related to the CH_4 + Ni(111) system is to extend full-dimensional (regarding the molecular degrees of freedom) quantum-dynamical treatments of reactions of molecules with surfaces to polyatomic molecules containing H atoms. Recent experiments on CH_4 + Ni demonstrate a dramatic dependence of the reaction probability on the initial vibrational state, in which the molecule is excited before the collision (*22*, *23*). Theoretical simulations are needed to explain these results and to predict which combined excitation of stretch and bend states should be most effective in promoting reaction. Such quantum-dynamics calculations can perhaps be done with the multiconfiguration time-dependent Hartree method (*24*). This method has recently been applied to photoinduced electron transfer in a dye-semiconductor system (*25*) in a calculation involving 39 vibrational modes of the dye molecule, a donor electronic state, and 1184 acceptor electronic states.

Theorists have also started to model electron-hole pair excitation in the scattering of H atoms from metal surfaces. In pioneering work, Trail *et al*. computed friction coefficients with time-dependent DFT (TDDFT) and used these coefficients in MDEF calculations on the scattering of H from Cu(111) (*26*). Derivations within the forced oscillator model implicit in MDEF, in which electron-hole pairs are treated as harmonic oscillators coupled linearly to nuclear motion,

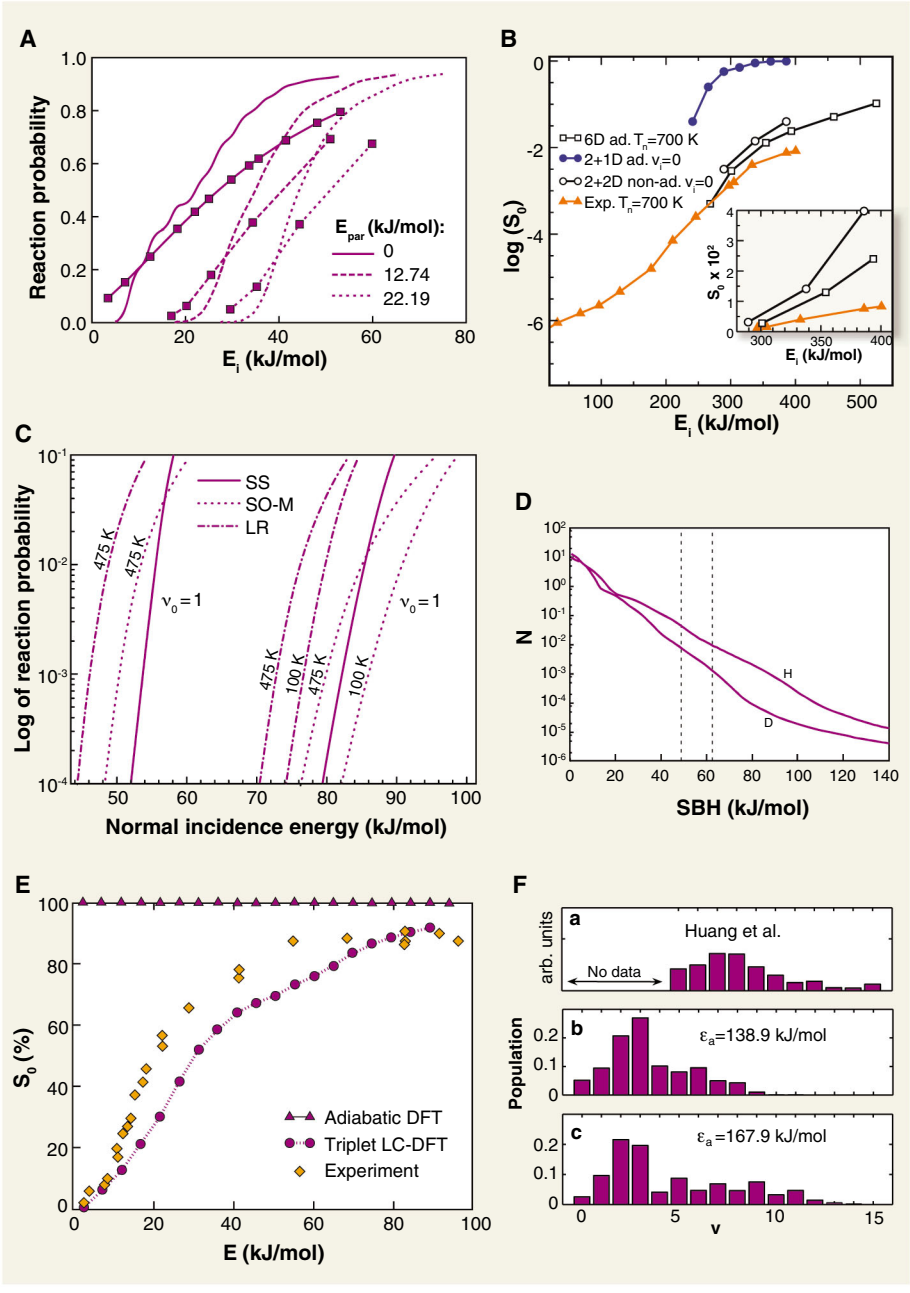


Fig. 2. (**A**) Theoretical reaction probabilities are compared with experimental results (squares) for H_2 + Pt(111) for normal and off-normal incidence with parallel translational incidence energy (E_{par}) as indicated. The results (*1*) are shown as a function of incident energy (E_i). (**B**) The probability of N_2 dissociation (S_0) on Ru(0001) is plotted as a function of normal incident energy, for 6D and reduced dimensionality calculations and for an experiment performed at a nozzle temperature (T_n) of 700 K. [Adapted from (*5*)] (**C**) Reaction probabilities computed for CH_4 + Ni(111) with the lattice-reconstruction (LR) model are compared with results of using the simple surface-oscillator model (SO-M) and that of using a static surface (SS) for two initial CH_4 vibrational states *v*. [Adapted from (*10*)] (**D**) The number of electrons per atom made available for detection over a Schottky barrier of a specific height (SBH) for H and D incident on the top site of Cu(111). The vertical lines span the range of Schottky barrier heights in experiments. [Adapted from (*26*)] (**E**) The dissociation probability of O_2 normally incident on Al(111) as a function of incident energy, from calculations using an adiabatic PES (triangles) and a diabatic PES (O_2 spin constrained to triplet) (circles) and from experimental data (diamonds). [Adapted from (*6*)] (**F**) The final vibrational-state population of NO scattering from Au(111) as obtained experimentally is compared with the computed final population for different values of the parameter denoting the anionic excited-state energy ε_a. [Adapted from (*3*)]

CREDIT: KATHARINE SUTLIFFE, *SCIENCE*

allowed the authors to calculate the number of excited electrons available for experimental detection (*12*) (Fig. 2D). This value and the large calculated H/D isotope effect were in good agreement with experimental data. Lindenblatt and Pehlke (*27*) used an ab initio molecular dynamics–spin-polarized TDDFT approach to compute electronic excitation spectra for H colliding with Al(111). Their work shows that, at a low incidence energy (5.8 kJ/mol) and on the first round trip through the chemisorption well, the greater part of the loss in the kinetic energy of H occurred through the coupling of the electron spin of H to the spins of the metal atoms. Their approach has so far been restricted to simple metals, but a "Newns-Anderson model" has recently been shown to also correctly describe dissipation to electron-hole pairs with nonadiabatic spin change; the approach provides accurate results for transition metals such as Cu and Ag (*28*).

MDEF calculations also addressed electron-hole pair excitation concurrent with dissociative chemisorption and its reverse process, associative desorption, of H_2 on/from Cu(111) and N_2 on/from Ru(0001) (*29*). Motion was modeled in the intramolecular and the molecule-surface distance coordinates. Although nonadiabatic effects were found to be minimal for H_2 + Cu(111), substantial energy (50 kJ/mol) was dissipated in the associative desorption of N_2 from Ru(0001). Energy dissipation to electron-hole pair excitation could explain why adiabatic calculations yield a much wider vibrational distribution of desorbed N_2 than experiments do (*18*) (fig. S3). An issue concerning MDEF calculations is that the dissipated energy scales quadratically with velocity and the velocities conjugate to the intramolecular distance, and the molecule-surface distance coordinates for which friction coefficients are high (*29*) will be lower if simulations model motion in all molecular degrees of freedom because of energy flow to the motion in the other simulated coordinates. Such high-dimensional MDEF simulations find little difference between adiabatic and nonadiabatic reaction probabilities for N_2 on W(110) (fig. S4) (*30*). However, this system shows a barrier in a geometry in which the friction coefficients are expected to be much smaller than at the N_2 + Ru(0001) barrier geometry.

Strong evidence exists that the dissociation reaction of O_2 on Al(111) proceeds nonadiabatically. Experimental data show an initially zero sticking probability increasing with collision energy, whereas adiabatic classical trajectory calculations based on a DFT PES yield unit reaction probability even at low energy (Fig. 2E) (*6*, *7*). A diabatic calculation performed using a PES obtained while constraining the spin state of O_2 to a triplet state was in good agreement with experimental data (Fig. 2E) (*6*). Similarly, good agreement with the sticking probability was obtained in wave-packet calculations on coupled diabatic PESs describing different successive charge-transfer states (*7*). Despite such good agreement with experiments, it is disturbing to note that the diabatic states used in the calculations correspond to different physics (spin conservation versus charge transfer).

Li and Guo (*2*) and Shenvi *et al.* (*3*) have recently studied multiquantum vibrational relaxation of highly vibrationally excited NO scattering from Pt and Au(111), respectively. Katz *et al.* (*31*) have simulated electron emission from a low-work-function metal surface in collisions with highly vibrationally excited NO. All these studies used diabatic potentials describing neutral NO and NO^- (after electron transfer from the surface), which were constructed on the basis of information from ab initio calculations, experimental data, and physical intuition. The calculations on multiquantum vibrational relaxation also included electron-hole pair excitation (*2*, *3*). The aim of these calculations was not yet to produce quantitative agreement with experimental data, but rather to see whether the model contains the essential physics to achieve qualitative agreement (see Fig. 2F for multiquantum vibrational relaxation) and to derive predictions for conditions under which experiments were not yet carried out (see fig. S5 for electron emission).

The greatest challenge facing theorists active in this field is to come up with an approach that can yield accurate predictions for molecule-metal surface reactions involving excited electronic states with potential curve crossings. Such an approach will also bring insight into the nature of excited states and their effect on reactions; for instance, it will show whether the activated character of O_2 dissociation on Al(111) (Fig. 2E) is due to spin quenching or charge transfer. One problem to be solved is that the so-called weak-coupling approximation inherent in the MDEF approach to electron-hole pair couplings (inherently present in scattering from metal surfaces) is unlikely to be applicable to curve crossings (*13*). Because the required potential energy surfaces and coupling connecting them can be obtained only from accurate electronic structure calculations (*32*), developing a method for computing accurate molecule-metal surface interaction energies for electronic ground states is a prerequisite for meeting the challenge set out above. Another major challenge is to come up with a good electronic structure approach to the description of excited molecule-metal surface states. As is the case for the ground state, embedding schemes involving high-level ab initio methods and DFT (*17*) may be of help, but because of the extended nature of metal electronic states, the existing schemes do not yet give a proper description of charge transfer (*7*). Finally, for systems involving molecules heavier than H_2, it will be necessary to accurately describe the effect of the phonons, and it will also be necessary to extend full-dimensional (regarding the molecular degrees of freedom) quantum-dynamical treatments to H atom–containing polyatomic molecules reacting with surfaces. As a result, for theorists, the simulation of molecule-surface reactions will remain an exciting playground for decades to come.

References and Notes

1. P. Nieto *et al.*, *Science* **312**, 86 (2006).
2. S. Li, H. Guo, *J. Chem. Phys.* **117**, 4499 (2002).
3. N. Shenvi, S. Roy, P. Parandekar, J. Tully, *J. Chem. Phys.* **125**, 154703 (2006).
4. A. Groß, A. Dianat, *Phys. Rev. Lett.* **98**, 206107 (2007).
5. C. Díaz *et al.*, *Phys. Rev. Lett.* **96**, 096102 (2006).
6. J. Behler, B. Delley, S. Lorenz, K. Reuter, M. Scheffler, *Phys. Rev. Lett.* **94**, 036104 (2005).
7. G. Katz, R. Kosloff, Y. Zeiri, *J. Chem. Phys.* **120**, 3931 (2004).
8. M. Head-Gordon, J. C. Tully, *J. Chem. Phys.* **103**, 10137 (1995).
9. G. J. Kroes, A. Gross, E. J. Baerends, M. Scheffler, D. A. McCormack, *Acc. Chem. Res.* **35**, 193 (2002).
10. S. Nave, B. Jackson, *Phys. Rev. Lett.* **98**, 173003 (2007).
11. Z. S. Wang, G. R. Darling, S. Holloway, *Phys. Rev. Lett.* **87**, 226102 (2001).
12. B. Gergen, H. Nienhaus, W. H. Weinberg, E. W. McFarland, *Science* **294**, 2521 (2001).
13. A. M. Wodtke, J. C. Tully, D. J. Auerbach, *Int. Rev. Phys. Chem.* **23**, 513 (2004).
14. J. D. White, J. Chen, D. Matsiev, D. J. Auerbach, A. M. Wodtke, *Nature* **433**, 503 (2005).
15. B. Hammer, M. Scheffler, K. W. Jacobsen, J. K. Nørskov, *Phys. Rev. Lett.* **73**, 1400 (1994).
16. Y. Zhao, D. G. Truhlar, *J. Chem. Phys.* **125**, 194101 (2006).
17. T. Klüner, N. Govind, Y. A. Wang, E. A. Carter, *J. Chem. Phys.* **116**, 42 (2002).
18. L. Diekhöner *et al.*, *J. Chem. Phys.* **117**, 5018 (2002).
19. H. F. Busnengo, W. Dong, P. Sautet, A. Salin, *Phys. Rev. Lett.* **87**, 127601 (2001).
20. E. Watts, G. O. Sitz, *J. Chem. Phys.* **111**, 9791 (1999).
21. M. Dohle, P. Saalfrank, T. Uzer, *Surf. Sci.* **409**, 37 (1998).
22. R. D. Beck *et al.*, *Science* **302**, 98 (2003).
23. L. B. F. Juurlink, R. R. Smith, D. R. Killelea, A. L. Utz, *Phys. Rev. Lett.* **94**, 208303 (2005).
24. M. H. Beck, A. Jäckle, G. A. Worth, H.-D. Meyer, *Phys. Rep.* **324**, 1 (2000).
25. I. Kondov, M. Cícek, C. Benesch, H. Wang, M. Thoss, *J. Phys. Chem. C* **111**, 11970 (2007).
26. J. R. Trail, D. M. Bird, M. Persson, S. Holloway, *J. Chem. Phys.* **119**, 4539 (2003).
27. M. Lindenblatt, E. Pehlke, *Phys. Rev. Lett.* **97**, 216101 (2006).
28. M. S. Mizielinski, D. M. Bird, M. Persson, S. Holloway, *Surf. Sci.* **602**, doi:10.1016/j.sus.2008.06.015 (2008).
29. A. C. Luntz, M. Persson, *J. Chem. Phys.* **123**, 074704 (2005).
30. J. I. Juaristi, M. Alducin, R. Diéz Muino, H. F. Busnengo, A. Salin, *Phys. Rev. Lett.* **100**, 116102 (2008).
31. G. Katz, Y. Zeiri, R. Kosloff, *J. Phys. Chem. B* **109**, 18876 (2005).
32. H. Köppel, W. Domcke, L. S. Cederbaum, *Adv. Chem. Phys.* **57**, 59 (1984).

33. The author thanks D. M. Bird and T. Klüner for useful discussions. This work was partly supported by Chemische Wetenschappen of the Nederlandse organisatie voor Wetenschappelijk Onderzoek through a PIONIER grant. The plots in Fig. 2, B to F, and figs. S2 to S5 were adapted with the consent and cooperation of the authors of the publications in which they originally appeared.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5890/794/DC1
Figs. S1 to S5
References

10.1126/science.1157717

PERSPECTIVE

Large-Scale Molecular Dynamics Simulations of Self-Assembling Systems

Michael L. Klein[1]* and Wataru Shinoda[2]*

Relentless increases in the size and performance of multiprocessor computers, coupled with new algorithms and methods, have led to novel applications of simulations across chemistry. This Perspective focuses on the use of classical molecular dynamics and so-called coarse-grain models to explore phenomena involving self-assembly in complex fluids and biological systems.

The proliferation of multiprocessor computers, combined with efficiently parallelized dynamics codes (*1*, *2*) and powerful visualization software (*3*), has heralded an explosive growth in applications of computation across chemistry. Now multiteraflop machines with thousands of processors can be accessed for extended periods of time. This in turn has led to classical molecular dynamics (MD) simulations for systems comprising a million atoms or so (*4*–*7*), which is sufficiently large to allow the components of Nature's machinery of life to be probed. Protein folding (*8*) and even molecular motors (*4*) are all yielding to the calculation power of such machines. Perhaps the best is yet to come, with the promise of petaflop machines in the near future (*9*) and the prospect of being able to use cheap high-performance graphics processors for MD studies (*10*, *11*).

These massive calculations are indeed successes for high-performance computing, notwithstanding the fact that the time scale of the MD trajectories on million-atom systems is typically less than 100 ns. However, in soft matter, the plethora of structures formed by self-assembling macromolecules typically involve micrometer-length scales embracing hundreds of millions of atoms, and phase transformations between structures occur on much longer time scales. Thus, such phenomena are currently inaccessible to brute-force classical MD treatment based on individual atoms. Also, in biological systems, phenomena such as lipid raft formation, and membrane fusion (*12*), system size, and especially the time scale, remain key issues.

Thus, a crucial question in carrying out large-scale MD simulations is, just how large a system does one need to access the phenomenon of interest? Which in the present context means, can one bridge scales ranging from tens of nanometers to micrometers? Even more important, how long should the MD simulation run, or can one bridge scales from tens of nanoseconds to milliseconds? To deal with these issues in soft matter and biosystems, so-called coarse grain (CG) models have been developed that allow MD to simulate at the mesoscale with system sizes approaching the micrometer range and phenomena taking place on time scales approaching milliseconds and beyond (*13*). In such models, the intermolecular potentials are either fitted to all-atoms (AA) MD simulations (*14*) and/or experimental data. In the case of atomistic models, the forces required to drive MD simulations typically are based on interatomic interaction potentials (*2*) derived in part from electronic structure calculations. Today's AA potentials can handle macromolecules, proteins, and DNA with remarkable success, especially given the inherent simplicity of the force fields that are mostly based on pairwise additive interaction potentials.

To illustrate some of the current possibilities for large-scale AA-MD and CG-MD computations in chemistry, we examine two examples. The first deals with the self-assembly of surfactant molecules and illustrates how structural transformations are now accessible to MD simulations, albeit with CG models. The second illustrates how a membrane protein can induce structural changes (curvature) in a biological membrane (*7*). In the latter case, both classical atomistic MD simulations and CG models have been used to great effect. For work that also has pushed the field to new frontiers, see (*18*–*20*).

Complex Fluids

The polymer community long ago adopted CG models with considerable success. Seminal studies by Smit *et al*. (*13*) inspired subsequent work on surfactants and lipids (*14*, *15*). The amphiphilic nature of surfactants leads to their aggregation and self-assembly into micelles, vesicles, and more complex morphologies under appropriate conditions. Understanding the role of molecular architecture in determining phase morphology and physical properties is a natural goal of MD simulations, which can then inform the design and synthesis of new molecules of interest to industry. However, even with generous access to multiteraflop machines, the spontaneous generation of

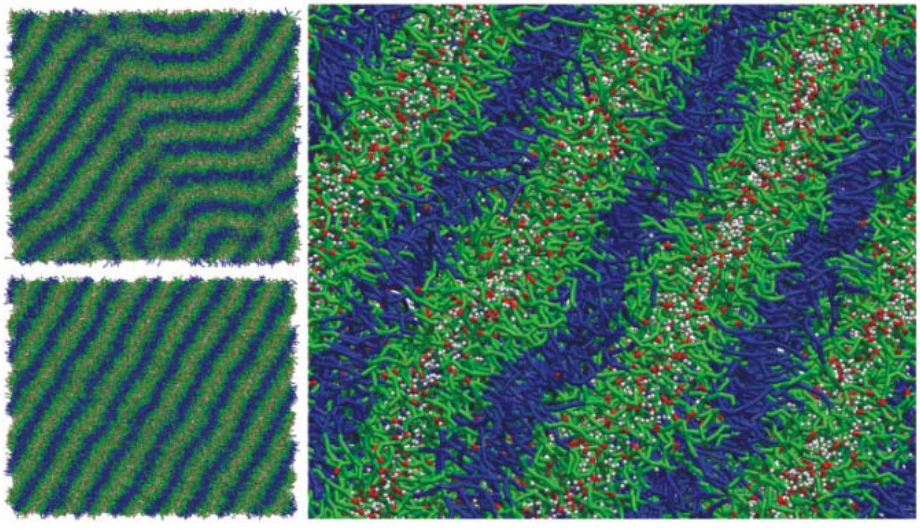


Fig. 1. Images taken from a MD simulation of an aqueous surfactant solution. The simulation system contains approximately 800,000 CG particles and represents a solution with 80% surfactant $C_{12}E_6$ and 20% water. The snapshots were taken after 100 ns (**top**) and ~500 ns (**bottom** and **right**) of MD simulation, starting from an appropriately hydrated hexagonal phase. The surfactant hydrophobic tails are shown in blue, the head groups in green and red, and the water in white. The phase transformation from initial hexagonal to final lamellar phase passes through a dislocated structure (top) before achieving the equilibrium structure (bottom). The image on the right shows expanded detail from the final structure. The CG parameters used in the MD simulations were taken from (*14*).

[1]Center for Molecular Modeling, Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104, USA. [2]Research Institute of Computational Sciences, National Institute of Advanced Industrial Science and Technology, 1-1-1 Umezono, Tsukuba 305-8568, Japan.

*To whom correspondence should be addressed. E-mail: klein@lrsm.upenn.edu (M.L.K.); w.shinoda@aist.go.jp (W.S.)

vesicles and wormlike structures starting from randomly dispersed surfactants in solution is highly challenging because of the time and length scales involved. The study of surfactant self-assembly into, for example, a cubic phase is typically beyond the capabilities of current computational resources if one desires an AA representation. Enhanced sampling techniques and CG representations of the surfactant and polymer molecules have therefore been developed.

With the present generation of CG models (*14*), it is possible to readily simulate systems containing over 1 million CG particles, corresponding to atomic systems of ~10 million atoms. Moreover, with a few thousand processors, it is possible to generate in a day MD trajectories that span about 100 ns. Such simulations are thus effectively approaching the regime relevant to laboratory experiments. A typical example of such a system, from our recent work (*14*), is shown in Fig. 1. The CG parameters used in the MD simulation are from (*14*). The system included more than 800,000 CG particles, composed of ~62,000 CG $C_{12}E_6$ molecules and more than 500,000 CG water particles (*14*). With an AA approach, this system would require nearly 5 million atoms to represent the solvent alone. The simulation was set up by starting with an equilibrated system containing 50 weight percent (wt %) $C_{12}E_6$ in the known hexagonal phase and slowly removing the water to obtain a system composed of 80 wt % $C_{12}E_6$. Fig. 1 shows images of the system as it evolves to the lamellar structure found in the experiment. After about 100 ns of MD simulation on the CG model, dislocation defects are clearly visible in the otherwise lamellar structure; at about 500 ns, the system has reached the lamellar phase, albeit with some minor defects still present. This CG model can also be applied to study Langmuir monloayers of surfactants at both the air/water and oil/water interfaces and the interaction of surfactants with nanoparticles and surfaces, to name just a couple of examples. Such CG computations are poised to become a predictive tool in surfactant design.

Biological Systems

A substantial beneficiary of multiteraflop machines has been chemical biology. Biological systems are intrinsically multiscale assemblies—micrometer-sized cells that contain integrated nanometer-scale functional units, such as protein oligomers. To date, the latter have been the main targets for MD. Successes have included folding studies on modest-sized proteins (*8*) and ion transport through membrane-bound protein bundles. Substantial efforts have been directed at simulating light-harvesting systems, molecular motors (*4*), and even a whole virus (*5*). Of particular note is the work on elucidating the mechanism for membrane curvature being induced by so-called Bin-amphiphysin-Rvs (BAR) proteins anchored to a membrane surface (*7*). Understanding this process is important because this phenomenon plays a role in cell membrane restructuring associated with endocytosis, the process whereby, for example, a virus is ingested by a cell.

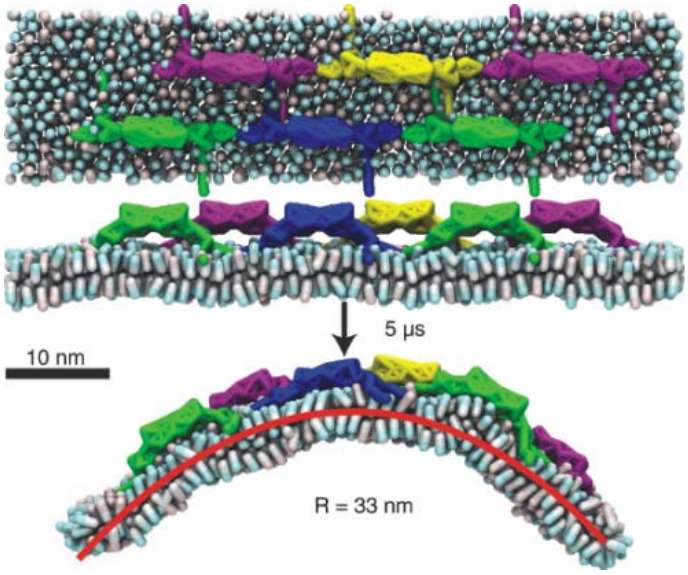


Fig. 2. Images taken from a MD simulation by Arkhipov *et al.*, (*17*) of biomembrane sculpting by protein-BAR domains. The simulation shown in the figure was carried out using a box with dimensions 100 by 16 by 50 nm and would correspond to a system of 10 million atoms in a fully atomistic representation. However, the computation was done using a shape-based CG model (*16*, *17*). As a result, there were 3265 CG beads in the system. Free space is needed around the membrane to allow for the bending. A similar simulation on a system four times larger and covering ~100 μs resulted in the spontaneous formation of a complete, stable membrane tube [private communication from the authors of (*16*)] with dimensions in agreement with experiment (*23*). *R* is the radius of curvature of the membrane induced by the protein-BAR domains.

MD simulations of biological systems comprising about 100 million atoms, such as that needed to probe endocytosis, are on the horizon (*9*), but the time-scale problem will remain the biggest impediment to real progress. As with surfactants and polymers, this has led to the promulgation of CG models, which reduce the number of interacting elements by "integrating out" selected groups of atoms via various schemes. Several approaches are showing substantial progress. These include, at one extreme, whole-protein building blocks in the case of the BAR proteins (*16*–*18*) and, at the other extreme, groups of about 10 atoms typically coalesced into a single CG bead (*14*). In this way, amino acid residues are typically reduced to one or two beads (*19*–*21*).

Such CG models affect computational demand in two ways. First, by reducing the number of interaction sites, the machine time required to compute the potential energy and forces for the system of interest is drastically reduced. Second, the resulting CG interaction "potentials" typically are much softer than in the AA case, thereby allowing a more ready exploration of phase space because the time step to integrate the MD equations of motion can be increased. Of course, this gain in efficiency comes with a reduction in the level of chemical detail. Successes with applying the CG approach to biosystems include rationalizing the mechanism of molecular motors (*4*) and first-structure determination of lipoprotein complexes used to transport lipids in vivo (*22*).

Figure 2 illustrates results of a MD simulation, recently carried out by Arkhipov *et al.* (*16*, *17*), of membrane sculpting by amphiphysin BAR-domain proteins. Specifically, six BAR domains (dark blue, green, yellow, and purple) were placed on top of a planar membrane patch consisting of zwitterionic dioleoylphosphatidylcholine lipids (light blue heads), with a 30% random mixture of dioleoylphosphatidylserine (DOPS) lipids (pink heads). The initial conformation of this system is shown viewed from the top and from the side. BAR domains scaffold the membrane with their concave, positively charged surface via electrostatic interactions with negatively charged DOPS lipids. A concerted action of BAR domains arranged in a lattice results in the development of a global membrane curvature on a time scale of several microseconds, with the resulting curvature radius of ~30 nm being close to that observed experimentally (*23*).

Outlook

This Perspective has focused on the exciting progress achieved through the use of large-scale classical MD to probe self-assembling macromolecular systems. We indicated how the use of CG models has helped to solve the problem of simulating micrometer-sized systems at millisecond time scales. However, we have not discussed the important area of multiscale modeling, which in principle reaches from the quantum mechanical (QM) world of atoms and electrons to the macroscopic world of finite element modeling (engineering). Over the years, multiscale methods have been developed to handle the problem of bridging length and time scales in chemistry. For example, in enzyme catalysis, approaches linking the QM description of the active site and the chemical reaction to the classical molecular mechanics (MM) of the protein have enabled the

study of function via so-called QM/MM methods. Again, the issue of time scales is relevant because the QM part of the system often dominates the computation. Unless approximate QM methods are employed, it becomes very challenging to run MD trajectories long enough to achieve adequate sampling, even with multiteraflop machines (*9*). Thus, new methods to enhance the sampling of rare events (such as the bond breaking in catalysis) are needed to help alleviate this problem (*24*). Another example of a multiscale method is the linking of electronic-structure methods to finite element-engineering approaches (*25*) to probe phenomena occurring at defects. Such methods are also likely to find application in the world of chemical biology.

With the predicted relentless increase in available computer resources, remaining issues that arise in the investigation of complex supramolecular organization and related phenomena in biological systems may be resolved via multiscale methods. However, for the near future, many aspects of the time-scale problem are likely to persist and remain beyond the scope of brute-force AA-MD simulations, even on petaflop machines. Thus, for the near term, CG and multiscale models will proliferate and yield exciting new data (*7*). The application of large-scale MD to chemical problems will be dominated by chemical biology, but the key to future progress will probably reside in new methods and algorithm development (*24*–*26*). Brute-force computation alone is going to be insufficient. Thus, the field is also poised to profit from future developments in the area of more refined methods for multiscale modeling to enable the efficient bridging between the QM, AA, CG, and finite element approaches. Finally, we have not discussed large-scale electronic structure calculations, which have been particularly successful in materials science, leading to the design of new catalysts (*27*). In the context of chemical biology, new functionals, and empirical corrections for van der Waals dispersion forces (*28*), are extending the range of validity of density functional theory to treat the interaction between large molecules such as DNA bases with high accuracy, which is likely to herald a new era for computational chemistry (*29*).

References and Notes

1. For example, see www.ks.uiuc.edu/Research/namd, http://lammps.sandia.gov, and www.gromacs.org.
2. For example, see www.charmm.org and http://amber.scripps.edu.
3. For example, see www.ks.uiuc.edu/Research/vmd.
4. A. Kitao *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4894 (2006).
5. P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, K. Schulten, *Structure* **14**, 437 (2006).
6. K. Y. Sanbonmatsu, C.-S. Tung, *J. Struct. Biol.* **157**, 470 (2007).
7. G. S. Ayton, P. D. Blood, G. A. Voth, *Biophys. J.* **92**, 3595 (2007).
8. B. Borrell, *Nature* **451**, 240 (2008).
9. J. Markoff, *New York Times*, 9 June 2008, www.nytimes.com/2008/06/09/technology/09petaflops.html?hp.
10. J. A. Anderson, C. D. Lorenz, A. Travesset, *J. Comput. Phys.* **227**, 5342 (2008).
11. J. A. van Meel, A. Arnold, D. Frenkel, S. F. Portegies Zwart, R. G. Belleman, *Mol. Simul.* **34**, 259 (2008).
12. H.-S. Yin *et al.*, *Nature* **439**, 38 (2006).
13. B. Smit *et al.*, *Nature* **348**, 624 (1990).
14. W. Shinoda, R. DeVane, M. L. Klein, *Mol. Simul.* **33**, 27 (2007).
15. J. C. Shelley *et al.*, *J. Phys. Chem. B* **105**, 4464 (2001).
16. A. Arkhipov, Y. Yin, K. Schulten, *Biophys. J.* doi:10.1529/biophysj.107.121160 (2008).
17. www.biophysj.org/cgi/rapidpdf/biophysj.108.132563v1
18. D. T. Mirijanian, G. A. Voth, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1204 (2008).
19. W. Treptow, S. J. Marrink, M. Tarek, *J. Phys. Chem. B* **112**, 3277 (2008).
20. M. Neri, M. Baaden, V. Carnevale, C. Anselmi, A. Maritan, P. Carloni, *Biophys. J.* **94**, 71 (2008).
21. L. Monticelli *et al.*, *J. Chem. Theory Comput.* **4**, 819 (2008).
22. A. Y. Shih, P. L. Freddolino, A. Arkhipov, K. Schulten, *J. Struct. Biol.* **157**, 579 (2007).
23. B. J. Peter *et al.*, *Science* **303**, 495 (2004).
24. E. Vanden-Eijnden, M. Heymann, *J. Chem. Phys.* **128**, 061103 (2008).
25. E. A. Carter, *Science* **321**, 800 (2008).
26. A. Barducci, G. Bussi, M. Parrinello, *Phys. Rev. Lett.* **100**, 020603 (2008).
27. F. Studt *et al.*, *Science* **320**, 1320 (2008).
28. V. R. Cooper *et al.*, *J. Am. Chem. Soc.* **130**, 1304 (2008).
29. M. Head-Gordon, E. Artacho, *Phys. Today* **61**, 58 (2008).

30. We thank A. Kohlmeyer, R. DeVane, and K. Anderson for discussion and K. Schulten, Y. Yin, and A. Arkhipov for providing Fig. 2. This research was supported in part by NSF (grant DMR 0520020) and NIH (grant GM 40712).

10.1126/science.1157834

PERSPECTIVE

Challenges in Modeling Materials Properties Without Experimental Input

Emily A. Carter

Simulations of materials behavior are an important component of materials science research, partly because measurements are indirect, requiring theoretical interpretation, and partly because often the ideal experiment simply cannot be performed (due to technological limitations). Empirical physical models used in this context often rely on parameters drawn from experiments on simpler systems, and so introduce various inaccuracies. In contrast, a quantum mechanical model can potentially offer an independent source of data more closely attuned to the complexities of the system at hand. This Perspective reviews current quantum mechanics–based materials modeling approaches and their successes and limitations, and offers a view to the future.

Imagine a day when we can invent new materials on demand just by entering into a computer the elements we want to use, the specific property we want to optimize, and our cost constraints. We are not there yet. Discovery and design of optimal, inexpensive materials will require both experimentalists and modelers working together to characterize the properties of new candidate materials. To this end, the theorist's role must be to develop and apply robust and accurate techniques that are first validated against well-characterized materials and then applied to new ones. This Perspective outlines the current state of quantum mechanics (QM)–based modeling of materials behavior and the frontier challenges that remain (*1*).

The first question to answer is, why focus on QM-based simulations? There is clearly much to be learned from atomistic, mesoscopic, and continuum-level modeling, but all such models require assumptions to be made about the physical laws governing the behavior of the material. Those assumptions can be good, but they are usually based on measurements taken from known, often simpler, materials, with unknown uncertainty as to their accuracy when applied to more complex materials. For instance, even one level less refined than QM, the classical mechanics or "atomistic" level, requires input of interatomic interaction potentials. Such potentials constrain the physics predicted via choice of functional form and fitted data, and do not exist for all combinations of elements. By contrast, in principle QM simply requires input of the atomic numbers of the elements, with no assumption made other than that the laws of QM hold.

The second question is, what properties can QM actually predict? We're told that QM should be able to predict everything, but in practice, each QM technique has its own set of limitations, so it is important to consider what can be predicted correctly qualitatively, quantitatively,

Department of Mechanical and Aerospace Engineering and Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544–5263, USA. E-mail: eac@princeton.edu

or not at all. Here's an undoubtedly incomplete list of material properties that can be obtained either directly from QM or by using the QM data in another theory: atomic-level structure of bulk crystals and interfaces; equations of state; phase diagrams; transition pressures; phase transition paths; melting points; elastic moduli; defect formation energies, including barriers to slip (dislocation nucleation); tensile and shear strengths at the nanoscale; fracture energies; phonon spectra; heat capacities; thermal expansion coefficients; thermal conductivity; electrical conductivity and conductance; magnetism; surface energies; and adsorption, desorption, diffusion, and reaction energetics, kinetics, and dynamics. Properties that QM cannot easily address are those that depend on complexities at larger length scales or integration over heterogeneities, such as plasticity in metals, and behavior of polycrystals or heterogeneous mixtures. For those, one must rely on multiscale modeling (*2*).

Unfortunately, no universal QM method exists that is appropriate for all materials and phenomena. Instead, we have a hierarchy of electronic structure techniques, all of which face a trade-off between accuracy and efficiency (Fig. 1). Ultimately, experimental validation of each theory is crucial, although it is difficult to design an experiment that measures exactly what was calculated (e.g., due to impurities and defects present in real samples).

We start with the most accurate: ab initio post-Hartree-Fock (HF) quantum chemistry (*3*) and quantum Monte Carlo (QMC) simulations (*4*). These methods incorporate the quantum mechanical effects of electron exchange (due to the Pauli exclusion principle) exactly and electron correlation (the motion of electrons is correlated to minimize repulsive forces) very accurately. No other physical approximations are made, both methods converge systematically toward exact solutions (within a fixed node approximation for QMC), and both provide properties of ground and excited electronic states. The fundamental difference between these two approaches is how observables are obtained. Ab initio post-HF quantum chemistry first solves for a many-electron wave function composed of the optimal linear combination of ground and excited state electron configurations expressed in terms of antisymmetrized products of orbitals, from which observables are then derived. QMC uses statistical sampling techniques to directly evaluate quantum mechanical observables, e.g., the energy. Ab initio quantum chemistry methods were developed initially to treat isolated molecules, whereas periodic QMC calculations on condensed matter came before applications to molecules. Three-dimensionally periodic post-HF methods are emerging, with second-order Møller-Plesset (MP2) perturbation theory as a first example (*5*). At present, other post-HF techniques have been used to study materials only via embedding schemes, in which a post-HF quantum chemistry calculation is performed on a cluster of atoms subjected to an embedding potential that accounts for the extended crystalline environment (*1*).

What are the main drawbacks of such methods? Foremost is computational expense. Post-HF quantum chemistry algorithms traditionally scale very poorly (where the required number of operations scale as N^5 or worse, for system size N) and QMC scales formally as $O(N^3)$ but with a large prefactor. Finite size effects turn out to be non-negligible in QMC, such that fairly large unit cells (at least tens of atoms) must be used to model a material, which adds to the cost. Encouragingly, the scaling of these methods has been reduced recently all the way to linear by the use of localized orbitals, local correlation, and integral screening techniques (*6*–*8*). On the other hand, these new methods calculate the wave function and energy but not forces on the nuclei in a linear scaling fashion. The latter is required to optimize structures and follow dynamics. Analytic gradients of post-HF energies have been available for decades, but such forces are very expensive. Only MP2 and various forms of multiconfigurational self-consistent field (MCSCF) forces are routinely used. QMC forces suffer not only from expense but also from the more fundamental problem of an infinite variance. However, recent clever renormalization and reweighting techniques have made accurate QMC forces available (*9*, *10*). A last challenge for QMC is treatment of transition metal–containing materials, which are problematic due to the highly oscillatory nature of the wave function; only a few examples of such calculations have appeared in the literature (*11*).

The more common family of methods for QM modeling of materials uses or builds upon density functional theory (DFT) (*12*), which is easily formulated with three-dimensional (3D) periodicity so that crystals can be simulated. It is much simpler and less costly than the methods above, solving either for a single-electron configuration in Kohn-Sham (KS) DFT (*13*) or, even simpler, solving for the density in orbital-free (OF) DFT (*14*). DFT is a formally exact ground state theory in which the material's energy is expressed as a functional of the electron density alone. However, two energy terms are usually approximated, because their exact density functional expressions remain unknown: the electron kinetic and exchange-correlation (XC) energies. Accurate electron kinetic energies require introduction of orbitals (*13*), which increases the algorithmic complexity from linear in OF-DFT to roughly cubic scaling in KS-DFT. However, for materials with a nonzero gap between occupied and unoccupied energy levels (the band gap), algorithms exist that recover linear scaling in KS-DFT above ~50 to 100 atoms (*15*). By contrast, OF-DFT can scale

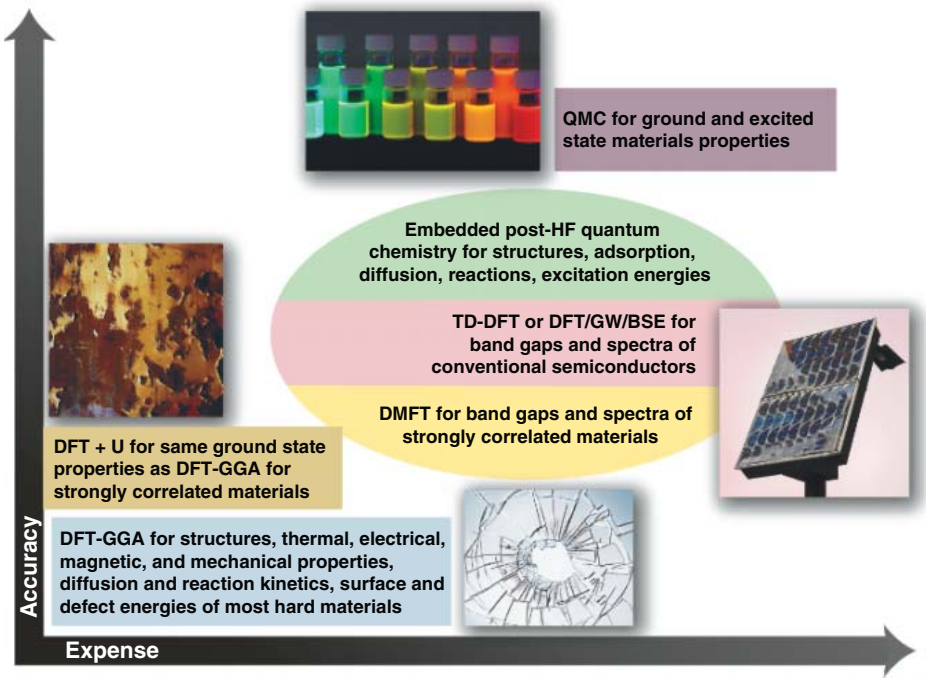


Fig. 1. Recommended quantum mechanics methods for various materials properties, distributed by accuracy and expense. The methods shown in the oval are comparable in cost and quality.

CREDITS: (CLOCKWISE TOP TO BOTTOM) EVIDENT TECHNOLOGIES, JUPITER IMAGES, DANIEL SMITH/ZEFA/CORBIS, ARTHUR S. AUBRY/PHOTODISC

linearly for all sizes but presently is limited by available kinetic energy functionals to providing accurate structural predictions only for main-group metallic materials, such as sodium, aluminum, or magnesium alloys (*14*).

The primary disadvantage of current incarnations of DFT is the approximate XC functional. In contrast to post-HF and QMC methods, no systematic means to improve the XC functional exists. Much effort has been expended to derive approximate XC functionals, either by keeping the functional ab initio (obeying known sum rules, bounds, and other physical constraints), or by allowing the functional to become (semi)empirical (fitting parameters in a chosen functional to experimental data). For most ground state properties, the generalized gradient approximation (GGA) XC functionals, which come in both ab initio and semi-empirical flavors, provide sufficient accuracy (Fig. 1) (*16*).

More costly hybrid XC DFT-GGA techniques include some exact HF exchange, which improves the description of any material with a band gap (*17*), but is not appropriate for metals, because HF band energies exhibit an unphysical logarithmic singularity at the Fermi level (the energy level separating occupied from unoccupied states). Because metals have occupied states around the Fermi level, this artifact manifests itself by producing fictitious charge and spin density waves in metals (*18*). Use of hybrid XC for metals, and attendant claims of accuracy, is thus a case of getting the right answer for the wrong reason. DFT+U is a more refined yet far less costly version of hybrid XC that also introduces exact X, but only for localized open-shell electrons for which the "self-interaction error" (due to approximate X) is greatest, instead of for all electrons as in hybrid XC. The delocalized electrons are treated with standard XC density functionals (*19*). DFT+U theory greatly improves the description of materials with this mixed type of electron distribution. This materials class includes mid-to-late first row transition metal oxides and sulfides, because the self-interaction error is largest for such multiply charged (i.e., +2 or higher) open-shell metal ions. DFT+U theory does require selection of one parameter, U-J, the difference between intra-atomic Coulomb and exchange energies. Recently, a means to systematically calculate U-J was proposed that uses exact exchange within ab initio HF theory, to render DFT+U theory non-empirical and free from any self-interaction error in the determination of U-J (*20*, *21*).

To calculate electronic or optical spectra of materials, one must have methods that can accurately predict excited states in condensed matter. Time-dependent DFT (TD-DFT) provides excitation energies, electron energy loss spectra, and absorption spectra, albeit within linear response theory (and therefore is limited to single-electron excitations). TD-DFT calculations on crystals have been rare, although new, more accurate XC functionals accounting for long-range interactions may encourage its use (*1*, *22*). Ionization energies and electron affinities, photoemission and inverse photoemission spectra, and band structures and band gaps can be obtained by using the many-body Green's function method within the GW approximation (*23*). The GW method takes DFT orbitals and energies as input and solves DFT-like equations in which the usual XC potential is replaced with the "self-energy," which is the product of a single-particle Green's function G and a dynamically screened Coulomb potential W. Neutral single-electron excitation energies and optical spectra can be obtained from the Bethe-Salpeter equation (BSE) (*23*), which takes DFT and GW data as input and accounts for electron-hole interactions. The GW/BSE and TD-DFT methods rely on ground state DFT being a good starting point. For strongly correlated electron materials such as first row late transition metal oxides, DFT with standard XC fails badly, and the GW approximation also breaks down. In this case, either QMC, embedded ab initio post-HF methods, or dynamical mean field theory (DMFT) (*24*) may be useful for predicting excitation energies, all of which should be able to describe strong electron correlations properly.

With these details in mind, I consider the appropriate methods to use in order to predict a given property for a given materials class as a function of accuracy and expense (Fig. 1). DFT-GGA is the workhorse, usually delivering qualitatively correct results at reasonable cost for chemical properties (molecular and crystal structures, bonding, molecular adsorption and diffusion on surfaces, bulk diffusion of atoms, and reaction energetics) of most hard materials, including metal alloys and semiconductors. Embedded ab initio post-HF methods correct qualitative and quantitative errors in structure and energetics engendered by DFT-GGA, but at additional complexity and expense (*1*). Therefore, DFT-GGA or embedded post-HF quantum chemistry would be the method of choice, e.g., for optimization of a heterogeneous catalyst. DFT-GGA also can estimate intrinsic mechanical properties [adhesion and fracture energies, tensile strengths when properly rescaled (*25*)] of brittle materials such as ceramics, as long as they are not strongly correlated materials like the transition metal oxides discussed earlier. For the latter, ab initio DFT+U theory may provide the best estimates of such chemical and mechanical properties at the nanoscale. Therefore, if kinetics of rust formation in a steel pipe or magnetic properties of Cr and Fe oxides for memory devices are of interest, ab initio DFT+U theory is a good choice.

Mechanical properties of metals will not be predicted properly by any QM method, as these properties are governed by plasticity, i.e., the motion of higher length-scale defects such as dislocations and grain boundaries. However, atomic-scale processes involved in the creation of these defects, such as barriers to slip (to create a dislocation), can be calculated reliably. Dislocation and grain boundary motion in main-group metals can be simulated reliably with OF-DFT, which gives insight into plastic deformation of simple metals (*26*). Thus, understanding how precipitates in an Al alloy affect its hardness or formability for use in a car door or airplane wing could be examined with OF-DFT.

Band gaps, photoemission and inverse photoemission spectra, and optical spectra are best calculated via GW/BSE for most semiconductors except those that are strongly correlated, for which DMFT is the current method of choice. TD-DFT and GW/BSE can be used to calculate qualitatively reasonable optical and electronic spectra of semiconducting nanostructures. Thus, the development of new infrared detectors or new solar cell materials might make use of these methods. Electrical conductivity and conductance, as well as thermal properties, can be calculated by using DFT as input; discussion of these methods is beyond the scope of this article.

Predicting the behavior of molecular, soft, amorphous, or heterogeneous materials poses notable challenges. Interactions in soft materials or molecular crystals are determined in large part by van der Waals forces, which are not described properly by standard XC functionals in DFT. Recent work applying specialized XC functionals (*27*), GW (*28*), and periodic local MP2 (*5*, *29*) have the correct physics and show great promise for treating such dispersion forces. With such techniques, properties of materials composed of, e.g., polymers, colloids, and proteins, will be accessible with QM.

Amorphous structures are difficult to model with QM for two reasons: The usual 3D periodic boundary conditions introduce correlation length artifacts, and it is never certain that a random amorphous structure generated, e.g., by quenching a dynamics trajectory, will be a representative one, due to the huge mismatch in time scales associated with molecular dynamics (MD) simulations versus actual laboratory quench times. Applying accelerated MD methods that bridge time scales may resolve the latter issue (*30*); the former is best addressed by the continued development of ever more efficient linear scaling QM methods to treat larger periodic unit cells. Heterogeneous mixtures of materials offer perhaps the most severe challenge for future materials modeling: The large numbers of constituent atoms require linear scaling algorithms and the presence of multiple types of materials may require different methods to be used simultaneously.

Simulation of complex materials motivates the recent strategy in materials simulation of mul-

tiscale modeling (*2*), which aims to bridge length and/or time scales to make overarching predictions of materials behavior, e.g., the mechanical response of a macroscopic-sized polycrystal subjected to shock or evaluating the thermal or electrical conductivity of a heterostructure. The methods include coupled QM-MD/MC and QM-(quasi)continuum methods, coupled atomistic-mesoscale techniques (e.g., MD coupled to dislocation dynamics/grain boundary motion), phase field/level set descriptions of microstructure, and continuum models informed by data passed from smaller length scales. Major unsolved issues in this area include how to transfer heat and mass across all scales; how best to bridge time scales and to coarsen or refine representations seamlessly and adaptively, without introducing additional physical assumptions; how to quantify uncertainties in this mixed representation; and ultimately, how to extract new concepts and new nonempirical physical laws from such simulations.

So where are we now, regarding modeling materials properties without experimental input? Admittedly, we still typically start with guidance from experiment regarding approximate initial structure (e.g., crystal symmetry, coordination number) and composition (e.g., elements, stoichiometry) before beginning to make predictions (*31*). But given such guidance, the strength of QM is that we can provide insight into how properties will change as we change the composition and structure, thereby furthering atomic-scale manipulation of the design of materials. We can suggest ways to improve catalysts for fuel production or coatings to inhibit corrosion and to protect against extreme heat. We can characterize optoelectronic, magnetic, and mechanical properties of materials important for designing the next solar cell, iPod, or fuel-efficient vehicle. Among the remaining challenges is how to move QM from characterizing a homogeneous material to doing the same for heterostructures: Real devices and macroscopic objects use multiple materials that may require multiple QM methods and higher length-scale theories operating in concert to predict an overall signal from a device, be it a mechanical, electrical, optical, or thermal signature. This challenge, of heterogeneity and integration of information, is the next frontier.

References and Notes

1. For a detailed technical review, see (*32*).
2. J. J. de Pablo, W. A. Curtin, Eds., "Multiscale Modeling in Advanced Materials Research," *MRS Bull.* **32** (2007).
3. A. Szabo, N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Dover, New York, 1989).
4. W. M. C. Foulkes, L. Mitas, R. J. Needs, G. Rajagopal, *Rev. Mod. Phys.* **73**, 33 (2001).
5. C. Pisani *et al*., *J. Chem. Phys.* **122**, 094113 (2005).
6. M. Schütz, H. J. Werner, *J. Chem. Phys.* **114**, 661 (2001).
7. T. S. Chwee, A. B. Szilva, R. Lindh, E. A. Carter, *J. Chem. Phys.* **128**, 224106 (2008).
8. A. J. Williamson, R. Q. Hood, J. C. Grossman, *Phys. Rev. Lett.* **87**, 246406 (2001).
9. M. W. Lee, S. V. Levchenko, A. M. Rappe, *Mol. Phys.* **105**, 2493 (2007).
10. C. Attaccalite, S. Sorella, *Phys. Rev. Lett.* **100**, 114501 (2008).
11. L. K. Wagner, *J. Phys. Condens. Matter* **19**, 343201 (2007).
12. P. Hohenberg, W. Kohn, *Phys. Rev.* **136**, B864 (1964).
13. W. Kohn, L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
14. Y. A. Wang, E. A. Carter, in *Theoretical Methods in Condensed Phase Chemistry*, vol. 5, *Progress in Theoretical Chemistry and Physics*, S. D. Schwartz, Ed. (Kluwer, Boston, 2000), pp. 117–184.
15. S. Goedecker, *Rev. Mod. Phys.* **71**, 1085 (1999).
16. S. Kurth, J. P. Perdew, P. Blaha, *Int. J. Quantum Chem.* **75**, 889 (1999).
17. J. Heyd, J. E. Peralta, G. E. Scuseria, *J. Chem. Phys.* **123**, 174101 (2005).
18. N. W. Ashcroft, N. D. Mermin, *Solid State Physics*, (Saunders, Fort Worth, TX, 1976).
19. A. I. Liechtenstein, V. I. Anisimov, J. Zaanen, *Phys. Rev. B* **52**, R5467 (1995).
20. N. Mosey, E. A. Carter, *Phys. Rev. B* **76**, 155123 (2007).
21. N. J. Mosey, P. Liao, E. A. Carter, *J. Chem. Phys.* **129**, 014103 (2008).
22. S. Botti, A. Schindlmayr, R. DelSole, L. Reining, *Rep. Prog. Phys.* **70**, 357 (2007).
23. G. Onida, L. Reining, A. Rubio, *Rev. Mod. Phys.* **74**, 601 (2002).
24. G. Kotliar *et al*., *Rev. Mod. Phys.* **78**, 865 (2006).
25. R. L. Hayes, M. Ortiz, E. A. Carter, *Phys. Rev. B* **69**, 172104 (2004).
26. S. C. Watson, P. A. Madden, *PhysChemComm* **1**, 1 (1998).
27. J. Kleis, B. I. Lundqvist, D. C. Langreth, E. Schroder, *Phys. Rev. B* **76**, 100201(R) (2007).
28. J. Harl, G. Kresse, *Phys. Rev. B* **77**, 045136 (2008).
29. D. Usvyat *et al*., *Phys. Rev. B* **76**, 075102 (2007).
30. A. F. Voter, F. Montalenti, T. C. Germann, *Annu. Rev. Mater. Res.* **32**, 321 (2002).
31. Fully unbiased QM predictions of structure with no experimental input are starting to emerge for simple materials. See e.g., (*33*).
32. P. Huang, E. A. Carter, *Annu. Rev. Phys. Chem.* **59**, 261 (2008).
33. C. J. Pickard, R. J. Needs, *Phys. Rev. Lett.* **97**, 045504 (2006).

34. Funding from the NSF, the Air Force Office of Scientific Research, the Office of Naval Research, the Army Research Office, the U.S. Department of Energy, and Princeton University is gratefully acknowledged. I also thank P. Huang for critically reading and discussing the manuscript.

10.1126/science.1158009

Latent Fingerprint Chemical Imaging by Mass Spectrometry

Demian R. Ifa, Nicholas E. Manicke, Allison L. Dill, R. Graham Cooks*

Latent fingerprints (LFPs) are distributions of exogenous and endogenous chemicals in a particular pattern. They potentially contain more forensic information than the simple identification of the subject; they may contain evidence of contact with explosives or substances of abuse. Chemically specific surface analysis methods are required to reveal the full information contained in LFPs. Such techniques often require derivatization (*1*) or addition of fluorescent or other tags (*2*). Mass spectrometry (MS) imaging (*3*) provides spatially specific chemical composition information on surfaces. Unlike other MS imaging methods, desorption electrospray ionization (DESI) MS (*4*) offers chemical imaging capabilities through mass spectra recorded without sample preparation from ordinary samples in their native environment.

In DESI, a solvent is electrosprayed onto the surface. Secondary scattered droplets are generated and evaporated, and their constituents mass analyzed. A mass spectrum is recorded in turn at each point on the surface by rastering the stream of charged droplets across the surface (*5*). This method has a spatial resolution of 150 μm. Spectroscopic techniques such as Fourier transform infrared (FTIR) and Raman also provide ambient imaging analysis and have better spatial resolution but poorer chemical specificity than MS techniques (*6*). However, the use of these methods to reconstruct fingerprint images from ordinary surfaces based on exogenous compounds has not been demonstrated (*6*).

We report the application of imaging MS for the chemical analysis of LFPs. Small amounts (5 μg) [an amount selected for its forensic relevance; see supporting online material (SOM) text] of drugs of abuse such as cocaine and Δ9-tetrahydrocannabinol (Δ9-THC, psychoactive component from cannabis) and explosives such as trinitrohexahydro-1,3,5-triazine (RDX, high-energy explosive) were imaged by DESI from LFPs produced after exposing the participant's fingers to solutions of these compounds (*7*) (see SOM text for further detail). In order to enhance the ionization of these compounds, additives such as ammonium hydroxide and sodium chloride were used in the solvent spray. The prints were recorded from ordinary surfaces such as glass, paper, and plastic. Figure 1A shows the distribution of cocaine (monitored as mass-to-charge ratio, *m/z*, 304) in a LFP on glass. The level of detail of the image, acquired with a pixel size of 150 μm by 150 μm, allows clear distinction of the ridges and minutiae. Figure 1C shows an ink fingerprint of the same individual blotted onto paper and optically scanned at a resolution of 600 points per inch (ppi). These images were used as input in fingerprint recognition software. The output images with automatically detected minutiae (Fig. 1, B and D) matched with use of the usual fingerprint identification tools (*7*). This experiment demonstrates that the mass spectrometry analysis allows physical identification of a subject. We recorded DESI images of LFPs for a wide range of chemicals and substrates, including tape-lifted fingerprints (fig. S1). Transfer of the original chemicals with retention of spatial relationships allowed the analytical mass spectrometer to be used to acquire chemical information on as well as the identification of a particular individual.

Endogenous components can also be imaged by using DESI. For instance, cis-hexadec-6-enoic acid, stearic acid, cis-octadec-8-enoic acid, palmitic acid, pentadecylic acid, myristic acid, and triacylglycerols were each used to record chemical fingerprints from sebum-rich LFPs.

The chemicals identified by their *m/z* ratios can be confirmed by their fragmentation patterns by using tandem mass spectrometry (MS/MS). For example, Δ9-THC and its nonpsychoactive isomer, cannabidiol, can be distinguished readily within a single LFP by MS/MS (*7*) (fig. S1). Distinguishing between overlapping fingerprints left by different individuals is extremely difficult optically. Imaging by mass spectrometry can readily distinguish each individual print with a distinctive exposure history to chemicals, such as Δ9-THC (fig. S1).

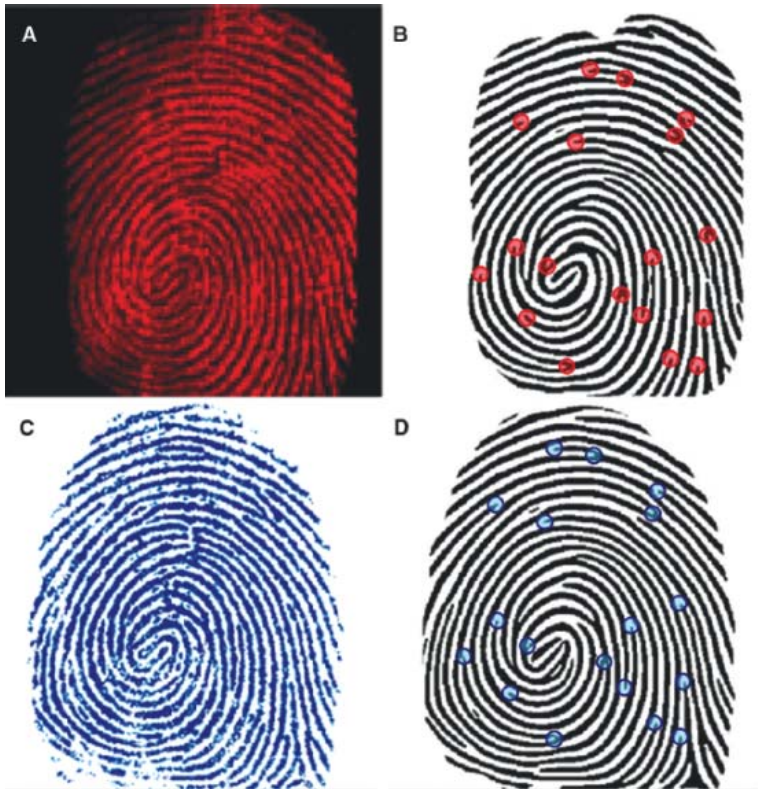


Fig. 1. (**A**) DESI image of distribution of cocaine on a LFP blotted on glass. (**B**) Computer-generated fingerprint from DESI image. (**C**) Ink fingerprint blotted on paper and optically scanned. (**D**) Computer-generated fingerprint from optical image. Some of the automatically detected points of interest (minutiae) are represented by dots in (B) and (D).

References and Notes

1. L. M. Dorozhkin, V. A. Nefedov, A. G. Sabelnikov, V. Sevastjanov, *Sens. Actuator B Chem.* **99**, 568 (2004).
2. R. Leggett, E. E. Lee-Smith, S. M. Jickells, D. A. Russell, *Angew. Chem. Int. Ed.* **46**, 4100 (2007).
3. L. A. McDonnell, R. M. A. Heeren, *Mass Spectrom. Rev.* **26**, 606 (2007).
4. Z. Takats, J. M. Wiseman, B. Gologan, R. G. Cooks, *Science* **306**, 471 (2004).
5. D. R. Ifa, J. M. Wiseman, Q. Y. Song, R. G. Cooks, *Int. J. Mass Spectrom.* **259**, 8 (2007).
6. M. Tahtouh, P. Despland, R. Shimmon, J. R. Kalman, B. J. Reedy, *J. Forensic Sci.* **52**, 1089 (2007).
7. Materials and methods are available on *Science* Online.

8. This work was supported by the Office of Naval Research (Research Tools program, grant N000140510454) and Prosolia Incorporated (Indianapolis, Indiana). It was performed at Bindley Bioscience Center at Purdue University, West Lafayette, Indiana, USA.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5890/805/DC1
Materials and Methods
SOM Text
Fig. S1
References

3 March 2008; accepted 27 May 2008
10.1126/science.1157199

Department of Chemistry and Center for Analytical Instrumentation Development, Purdue University, West Lafayette, IN 47907, USA.

*To whom correspondence should be addressed. E-mail: cooks@purdue.edu

The Rupture and Repair of Cooperation in Borderline Personality Disorder

Brooks King-Casas,[1,2] Carla Sharp,[2] Laura Lomax-Bream,[2] Terry Lohrenz,[1] Peter Fonagy,[2,3,4] P. Read Montague[1,2]*

To sustain or repair cooperation during a social exchange, adaptive creatures must understand social gestures and the consequences when shared expectations about fair exchange are violated by accident or intent. We recruited 55 individuals afflicted with borderline personality disorder (BPD) to play a multiround economic exchange game with healthy partners. Behaviorally, individuals with BPD showed a profound incapacity to maintain cooperation, and were impaired in their ability to repair broken cooperation on the basis of a quantitative measure of coaxing. Neurally, activity in the anterior insula, a region known to respond to norm violations across affective, interoceptive, economic, and social dimensions, strongly differentiated healthy participants from individuals with BPD. Healthy subjects showed a strong linear relation between anterior insula response and both magnitude of monetary offer received from their partner (input) and the amount of money repaid to their partner (output). In stark contrast, activity in the anterior insula of BPD participants was related only to the magnitude of repayment sent back to their partner (output), not to the magnitude of offers received (input). These neural and behavioral data suggest that norms used in perception of social gestures are pathologically perturbed or missing altogether among individuals with BPD. This game-theoretic approach to psychopathology may open doors to new ways of characterizing and studying a range of mental illnesses.

Theoretical and empirical work on cooperation has made tremendous strides by using game theory to provide a normative mathematical setting for understanding social exchange (*1*–*5*). This same game-theoretic framework has exposed some of the basic computations underlying social interaction and identified neural correlates of cooperation, reciprocity, and social signaling (*6*–*14*). Collectively, this work raises the possibility of a new approach to characterizing and understanding psychopathology from a normative perspective. In conditions ranging from psychosis to developmental and personality disorders, afflicted individuals often display a dramatically perturbed capacity to model others and to sense and respond appropriately to the social signals they emit (*15*–*17*). Consequently, using normative game-theoretic probes of social signaling in identified psychopathologies offers the opportunity to understand some of the components of these disorders in terms of malfunctioning computations (*18*, *19*).

We examined two groups of individuals that vary in their capacity to maintain stable interpersonal relationships and used a multiround economic exchange game (Fig. 1) (*20*) to probe two fundamental components underlying the capacity to sustain a successful social exchange: cooperation and its repair. Specifically, we studied a group of control individuals and a second group of individuals diagnosed with borderline personality disorder (BPD), a psychiatric disorder characterized by unstable relationships, affective dysregulation, and a broad incapacity to trust appropriately the actions and (possible) motives of others (*21*–*23*).

Successful cooperation between two agents requires a range of intact computations including the capacity to sense, value, and respond to social gestures exchanged with one's partner (*24*–*26*). However, cooperative exchange is fragile and can easily be ruptured by accident (e.g., through impoverished models of social partners) or intent (*27*). The successful repair of broken cooperation in an iterated exchange requires the capacity to coax one's partner back into cooperation through the medium of generous gestures (*28*–*32*). In an economic exchange game, such gestures are encoded as money units, which exposes their immediate cost to each subject. However, for such gestures to be meaningful, cooperating individuals must possess norms that accurately encode what is expected of themselves and their partners, and they must sense when these norms have been significantly violated (*33*). We pursued the hypothesis that social exchange norms of subjects with BPD are either pathologically perturbed or missing altogether, preventing sustained cooperation or the repair of cooperation after it breaks down.

Before scanning, all participants (healthy subject group and BPD group) underwent diagnostic and symptom assessment and were matched on demographic variables including sex, age, education, and verbal IQ (see table S1) (*20*). Within each trust exchange, cooperation occurs when an investor and trustee act in a manner that mutually benefits both players. For example, if an investor sends $20 to a trustee, and the trustee splits the tripled investment ($60) with the investor, both the investor and trustee profit. The investor earns $10 more, and the trustee earns $30 more, than if the investor had sent

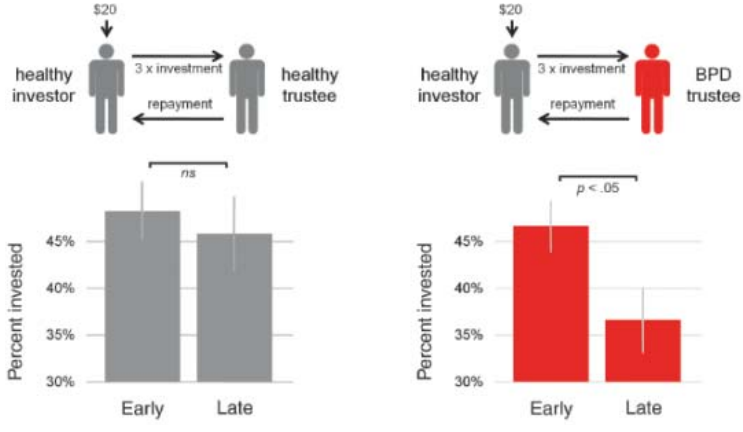


Fig. 1. Cooperation fails across rounds when individuals with BPD engage in a repeated exchange of trust. Among 38 healthy investors (gray) paired with healthy trustees (gray), investments were large and sustained across early (1 to 5) rounds and late (6 to 10) rounds of the game. However, among 55 healthy investors (gray) paired with 55 trustees with BPD (red), a decrease in investment level from early to late rounds of the game indicates a failure in cooperation across the iterated exchange. Mean percent invested and SEM are plotted.

[1]Computational Psychiatry Unit and Department of Neuroscience, Baylor College of Medicine, 1 Baylor Plaza, Houston, TX 77030, USA. [2]Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, 1 Baylor Plaza, Houston, TX 77030, USA. [3]Research Department of Clinical, Educational, and Health Psychology, University College London, Gower Street, London WC1E 6BT, UK. [4]Anna Freud Centre, London NW3 5SD, UK.

*To whom correspondence should be addressed. E-mail: rmontague@hnl.bcm.edu

nothing. However, if a trustee does not repay at least the amount invested, the investor accrues no benefit from the exchange, likely triggering smaller subsequent investments. Thus, increased cooperation is seen with increased money exchanged across the course of the 10-round game. In the current study, subjects diagnosed with BPD played the trustee role against a healthy investor; as expected, each game began normally with the healthy investor showing a level of trust comparable to that of healthy investors playing healthy trustees (Fig. 1).

BPD subjects send signals that break cooperation. In early rounds of this multiround game, investment levels did not differ between subjects partnered with BPD trustees and subjects partnered with healthy trustees. However, in late rounds of the game, investments were significantly lower for dyads with a BPD trustee relative to control dyads with a healthy trustee (Fig. 1). This definitive downward shift in investment levels for dyads with BPD trustees reflects a break in cooperation [see figs. S2 to S4 and (*34*) for additional detailed analyses]. We strongly suspected that this breakdown was caused by the social signals emitted by the BPD trustee and were not due to a sampling of healthy investors that happened to be uncooperative. To test this claim empirically, we recruited an additional cohort of 68 healthy investors to play an adaptive computer agent designed to play as either a BPD trustee or a healthy trustee. A significant within-subject effect confirmed that BPD trustee behavior alone elicits a breakdown in cooperation (see fig. S5) (*20*).

Coaxing pays into the future. Although the BPD subjects showed a consistent tendency to rupture cooperation with their healthy partners, the issue of broken cooperation arises during all interpersonal exchange. Individuals bring to any two-party interaction variability in their expectations of partner behavior, variable sensitivities to deviations from these norms, and variable responses to these deviations. These issues highlight a feature that plagues all social exchange—breakdowns in cooperation are common, and successful social agents must sense (or even anticipate) these failures and select actions to repair them (*28*–*32*). Consequently, we specifically sought to identify strategic responses to cooperation breakdowns that allow healthy trustees to maintain high levels of cooperation into the later rounds of the exchange. We focused on rounds in which cooperation was low, that is, rounds in which investments of $5 or less were made. Low investments act as a viable proxy for broken cooperation because in this economic exchange game and other related games (*7*, *9*, *10*, *35*), opening investments to a responder (here the trustee) are typically large. For investments to become small, a responder has to "prove" that they are not worthy of a large investment by acting in an untrustworthy manner. Such ruptures of cooperation, nevertheless, provide the trustee an opportunity to repair the broken cooperation. If trustees repay a large fraction of the tripled investment, they signal their trustworthiness in the presence of low offers and may garner greater investments on subsequent rounds (Fig. 2A). We term this behavior "coaxing" and show in Fig. 2B that coaxing pays dividends for four rounds into the future. In this same figure, we summarize the results for low offers when the trustees do not coax—trust and cooperation remain broken. So in the context of this multiround game, coaxing repairs cooperation and pays off generously for the coaxer. Remarkably, comparison of healthy trustees to BPD trustees found healthy players to be almost twice as likely as BPD players to coax in the presence of low offers (i.e., repay a third or more of the investment) (Fig. 2C and fig. S3).

BPD subjects have perturbed norm deviation responses in insular cortex. These behavioral data with humans and computer agents suggest strongly that BPD subjects are likely to emit responses that cause cooperation to rupture, and, once ruptured, they show substantially diminished rates of coaxing (generous gestures) to repair the break (Fig. 2C). A question naturally arises about whether there is a comprehensible and consistent neural correlate of the rupture in cooperation that serves to cue the repair (or not) of cooperation. To investigate the possible neural origins of failures in cooperation and the lack of coaxing, we first sought between-group differences in the sensitivity of trustee brains to the revelation of small relative to large investments. A general linear model (GLM) analysis identified several regions that significantly differentiated the controls from the BPD individuals, including bilateral anterior insula, along with frontal and parietal areas (Fig. 3, top left; also see table S3 and fig. S6). Subsequent within-subject analyses identified only a single region in controls with greater response to small, relative to large investments—bilateral anterior insula. In particular, both a GLM analysis (Fig. 3, top middle; also see table S4) and region-of-interest (ROI) analysis of voxels identified in the between-group comparison show a strong

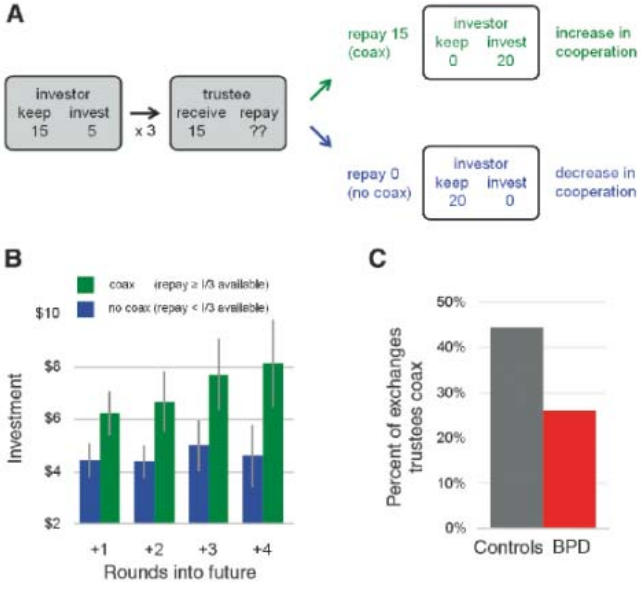


Fig. 2. (**A**) Coaxing by trustees engenders trust from investors. Cooperation within the iterated trust game is expressed as high levels of investment combined with equitable sharing of investment returns (see fig. S2 for normative behavior). When cooperation falters (i.e., when investments are low), trustees can rebuild trust and cooperation by coaxing back higher and higher investment levels. In this schematic representation, the investor entrusts only a small portion of their endowment ($5 of $20 points) to their partner. The $15 received by the trustee is far less than the $60 that would have been received if the investor had sent the entire endowment. Thus, in order to maximize their gains, a trustee must induce cooperation in a repeated exchange. The trustee can coax back trust from their partner by repaying a large proportion of the tripled investment–signaling their own trustworthiness and, thus, eliciting larger subsequent investments. Conversely, if the trustee does not coax and instead keeps a large proportion of the tripled investment, the investor is likely to invest less on subsequent rounds and cooperation will continue to devolve. (**B**) Coaxing pays off. The effect of such coaxing is summarized in the effect of repayment on investments in subsequent rounds among healthy subjects. This analysis is restricted to low investments (all rounds with investments <$5). Blue bars depict mean ± SE investment level following small repayments (repayment less than one-third of the tripled investment; "no coax" condition); green bars depict investment level following large repayments (more than or equal to one-third; "coax" condition). The analysis indicates that coaxing elicits larger investments in rounds following coaxing behavior (green) than for rounds following no coaxing (blue). Note that the coaxing-related increase in investment persists across rounds, such that a large repayment in the current round earns larger investment up to four rounds into the future. (**C**) BPD subjects coax less. Healthy trustees are twice as likely as BPD trustees to coax when cooperation between players is low. Specifically, healthy trustees are more likely to make a large repayment (≥ investment amount) after having received a small investment (≤$5). Conversely, BPD trustees are more likely to make a small repayment (< investment amount) after receiving a small investment. The *y* axis indicates the proportion of exchanges that trustees repay more than or equal to the investment amount after receiving a small investment (≤$5).

negative relation between bilateral anterior insula activity and investment level (Fig. 3, bottom middle). In stark contrast, in BPD subjects comparable GLM and ROI analyses (Fig. 3, top right and bottom right, respectively; also see table S5) showed no systematic relation of insula response and investment level. Additional ROI analyses of medicated and unmedicated BPD participants (Fig. 4, left and middle), and BPD participants matched on income level (Fig. 4, right) to healthy trustees confirm that the diminished response cannot be attributed to group differences in sex, age, verbal IQ, education, income level, or medication status.

Repayment-related insula activity intact in BPD subjects. In earlier work using this same multiround trust game (*35*) and single-shot ultimatum games (*12*), small offers increase the probability of defection in the partner, and such behavior is preceded by increased activity in the anterior insular cortex [also see (*36*) for review]. Consequently, we pursued an additional region-of-interest analysis of the anterior insular cortex to examine whether activity in this region predicts small repayments by trustees (Fig. 5). Using the same voxels originally identified in the group by investment-level GLM, we found that both healthy and BPD trustees show a negative relation between insula activity and repayment level during trustee decisions. That is, despite the diminished insula response exhibited by BPD trustees during the receipt of small offers (input), a robust insula response in this group is observed preceding decisions to repay small amounts (output). Thus, healthy subjects show a strong linear relation between anterior insula response and both the monetary offer sent from their partner [input (Fig. 3)] and their monetary response back to their partner [output (Fig. 5)], whereas BPD subjects' insula yielded a linear relationship only to their response back to their partner [output (Fig. 5)], but showed no differential activity upon receipt of offers from their partner (Fig. 3). Moreover, the BPD insula response, on receipt of offers from the investor, showed the same average level of activity across all investment levels, a fact true for both medicated and unmedicated subjects (Fig. 4).

Although the insular cortex has traditionally been associated with pain perception, representation, and interoception (*37*, *38*), responses in the anterior insula have recently been identified in social interactions where norms are violated (*12*, *14*, *39*, *40*). In an ultimatum game, in which one player offers to split an endowment with a second player, small offers are perceived as unfair and typically refused (*41*–*43*). In such instances, anterior insula activity both scales negatively with offer size and predicts whether the offer is subsequently rejected (*12*). In a related finding, anterior insula activity of observers is greater when a punishment is applied to players perceived as fair relative to players perceived as unfair (*13*). In nonsocial tasks, activity in the anterior insula has been found to encode representations of risk and uncertainty about decision outcomes (*44*–*50*). The association of the insula with a representation of outcome variance suggests that the insula may encode the distribution of likely outcomes in social interactions; that is, responses in the anterior insula may indicate social norm violations within interpersonal contexts. Furthermore, violations of such norms could serve as control signals that update expectations about social partners or at least inform learning about a subset of parameters associated with one's partner. This possibility is supported by the work of Preuschoff and Bossaerts, who have recently reported prediction errors of risk to evoke strong responses in the bilateral anterior insula, consistent with this speculation (*49*–*51*). Taken together, these data suggest that the strong negative relation between offer size and activity in the anterior insula seen in healthy trustees reflects the perceived violation of social norms in the two-party trust exchange.

Moreover, in BPD, the robust insula response preceding one's own repayments is discordant with the absent insula relation to the investments

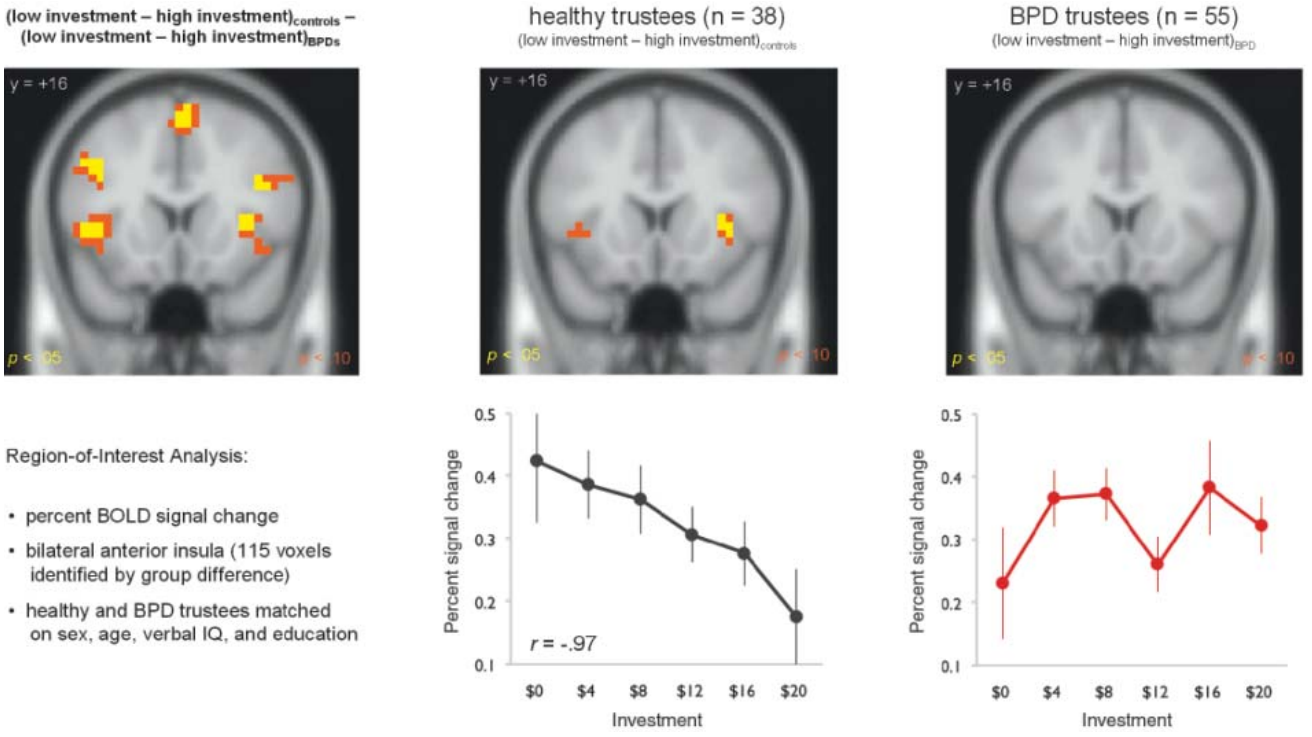


Fig. 3. Response of 38 healthy trustee brains and 55 BPD trustee brains to level of cooperation. (**Top**) Results of both between- and within-group GLM analyses identified cortical regions with greater response to small investments ($I \leq \$5$) relative to large investments ($I > \$10$). All significant voxels are conservatively corrected using false-discovery rate procedures to the $P < 0.05$ level in yellow; $P < 0.10$ in orange (see tables S3 to S5). (**Bottom**) Percent change in hemodynamic signal was averaged from the 115 most significant voxels identified in the group-level GLM (top left and table S3) during the 4- to 8-s period following the revelation of investment. The means ± SE of the resulting signal are plotted in $4 bins. Responses in bilateral anterior insula in healthy trustees scale strongly and negatively with the size of investment ($r = -0.97$; bottom middle). In contrast, similar analyses in individuals with BPD showed no such relation.

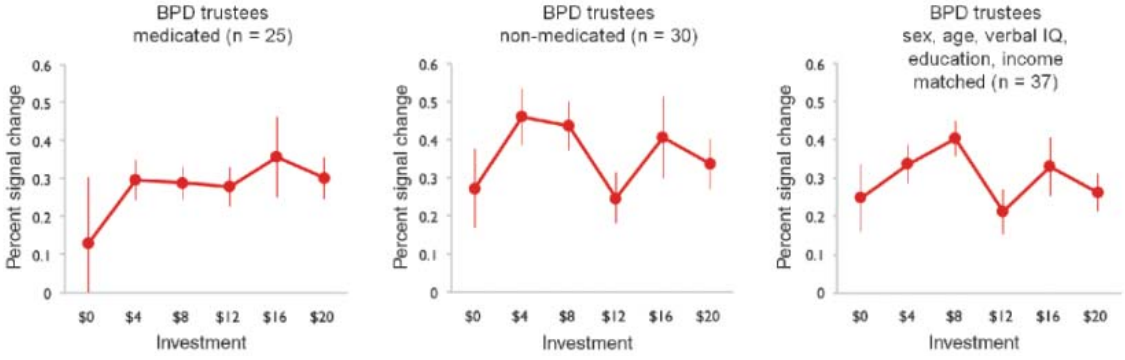


Fig. 4. Diminished relation of insula activity to investment level is not attributable to medication status or income level in BPD trustees. ROI analyses within subsamples of BPD trustees were performed as described in Fig. 3. The systematic negative relation of investment level to bilateral anterior insula activity seen in healthy controls is absent in both medicated and unmedicated BPD trustees, as well as in BPD trustees matched to the income level of the control group.

of others and highlights the specificity of the observed deficit in BPD. That is, in BPD, the apparent insensitivity of the insula only to offer level size, and not their own repayment, suggests two possibilities: (i) Monetary reward is not reinforcing to individuals with BPD; or, (ii) low offers are not perceived to be a violation of social norms to individuals with BPD. Previous work has shown monetary reward to be a potent reinforcer in natural and laboratory settings in this group (*52*), which makes the former less likely. Furthermore, studies of interpersonal and emotional processing in BPD suggest that this group holds negative expectations of social partners and exhibits negative evaluative biases (*53*, *54*), consistent with the presence of atypical social norms.

Economic probe corresponds with self-report measure of trust. To compare social norms of healthy trustees to BPD trustees within the current study, we also implemented a self-report measure of trust (*55*) that gauges expectations across a variety of social situations and with a variety of social agents. Individuals with BPD expressed significantly lower levels of self-reported trust relative to healthy controls [$P < 0.005$ (Fig. 6)], a finding that agrees with the diminished trust exhibited by this group in the economic exchange. Together, these results suggest that the diminished insula response to low offers does indeed reflect atypical social norms in this group. Put another way, the low offers from social partners do not violate the social expectations of individuals with BPD, which accounts for the diminished insula response in the BPD group.

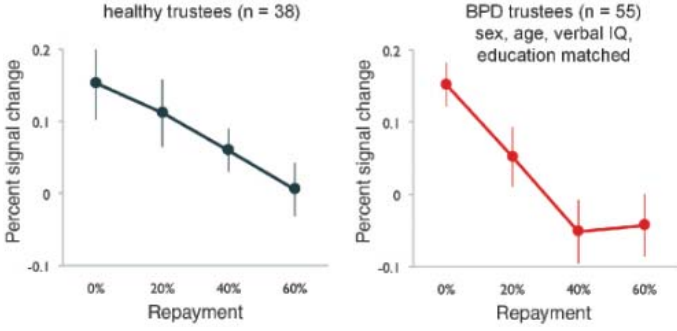


Fig. 5. Both healthy and BPD trustees exhibit a systematic negative relation of repayment level and bilateral anterior insula activity during their decision to repay. ROI analyses were performed using bilateral anterior insula voxels identified in the between-group analysis illustrated in Fig. 3 and used in the ROI analyses of Figs. 3 and 4. To investigate whether the insula activity also scales with trustee responses to norm violations, insula activity during trustee decisions is plotted by repayment level. Both healthy and BPD trustees show a strong negative relation between activity and repayment level. As trustees repaid more than 60% of the tripled investment in only ~ 7% of rounds, high repayment rounds are excluded from this analysis.

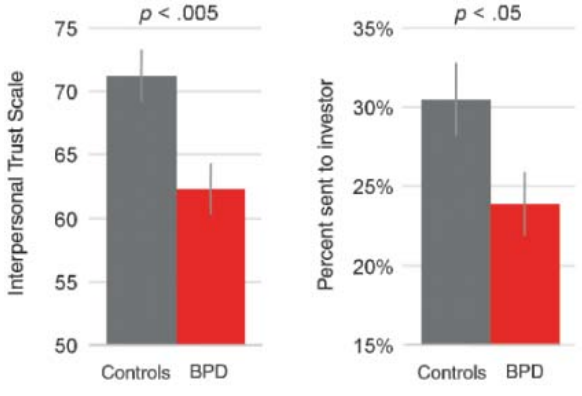


Fig. 6. Behavioral and self-report levels of trust indicate perturbed social norms among individuals with BPD. (**Left**) Individuals with BPD report lower trait levels of trust using a self-report measure [Interpersonal Trust Scale (*55*)]. (**Right**) Individuals with BPD repay less than the healthy group, which indicate lower levels of trust within the exchange game (also see fig. S2).

The atypical social norms of individuals with BPD have considerable implications for both partners in a repeated cooperative exchange. Specifically, the incapacity of individuals with BPD to recognize norm violations as such and to respond so as to signal trustworthiness to their partner can result in an inability of both individuals to effectively model the intentions of one another. For example, an investor might send a small amount to a trustee to signal lack of trust and may do so with the expectation that the trustee will, in turn, signal a desire to cooperate through a large repayment (coaxing). However, the lack of expected response (failure to coax) may inadvertently signal additional lack of cooperation, regardless of the trustee's intent. In short, the atypical social norms of a single partner can have deleterious effects on the ability of both players to model and predict the actions of social partners and, thus, to inhibit both players' ability to mutually benefit from cooperative exchange.

A number of studies have utilized resting-state and emotional challenge paradigms to investigate personality disorders; however, the current study is the first large-scale investigation of interpersonal behavior among individuals diagnosed with a personality disorder (Axis-II) (*22*) using a normative economic exchange game. The critical role that interpersonal deficits play across a variety of such disorders, along with the substantive contribution that neuroimaging studies have made in elucidating the etiology of Axis-I clinical disorders, recommend the future use of game-theoretic probes of interpersonal interaction to uncover the neural processes underlying social pathologies (*18*, *19*, *56*).

References and Notes

1. R. Axelrod, W. D. Hamilton, *Science* **211**, 1390 (1981).
2. H. Gintis, *Game Theory Evolving* (Princeton Univ. Press, Princeton, NJ, 2000).
3. M. A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life* (Harvard Univ. Press, Cambridge, MA, 2006).
4. R. L. Trivers, *Q. Rev. Biol.* **46**, 35 (1971).

5. J. von Neumann, O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton Univ. Press, Princeton, NJ, ed. 2, 1947).
6. D. J. de Quervain *et al.*, *Science* **305**, 1254 (2004).
7. M. R. Delgado, R. H. Frank, E. A. Phelps, *Nat. Neurosci.* **8**, 1611 (2005).
8. D. Knoch, A. Pascual-Leone, K. Meyer, V. Treyer, E. Fehr, *Science* **314**, 829 (2006).
9. K. McCabe, D. Houser, L. Ryan, V. Smith, T. Trouard, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11832 (2001).
10. F. Krueger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20084 (2007).
11. J. Rilling *et al.*, *Neuron* **35**, 395 (2002).
12. A. G. Sanfey, J. K. Rilling, J. A. Aronson, L. E. Nystrom, J. D. Cohen, *Science* **300**, 1755 (2003).
13. T. Singer *et al.*, *Nature* **439**, 466 (2006).
14. M. Spitzer, U. Fischbacher, B. Herrnberger, G. Grön, E. Fehr, *Neuron* **56**, 185 (2007).
15. M. Sprong, P. Schothorst, E. Vos, J. Hox, H. Van Engeland, *Br. J. Psychiatry* **191**, 5 (2007).
16. M. Brüne, U. Brüne-Cohrs, *Neurosci. Biobehav. Rev.* **30**, 437 (2006).
17. J. Rogers, E. Viding, R. J. Blair, U. Frith, F. Happé, *Psychol. Med.* **36**, 1789 (2006).
18. P. H. Chiu *et al.*, *Neuron* **57**, 463 (2008).
19. J. K. Rilling *et al.*, *Biol. Psychiatry* **61**, 1260 (2007).
20. Materials and methods are available as supporting material on *Science* Online.
21. K. Lieb, M. C. Zanarini, C. Schmahl, M. M. Linehan, M. Bohus, *Lancet* **364**, 453 (2004).
22. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR* (American Psychiatric Association, Washington, DC, ed. 4, text revised, 2000).
23. P. Fonagy, A. W. Bateman, *J. Clin. Psychol.* **62**, 411 (2006).
24. C. D. Frith, U. Frith, *Brain Res.* **1079**, 36 (2006).
25. T. Singer, *Neurosci. Biobehav. Rev.* **30**, 855 (2006).
26. A. G. Sanfey, G. Loewenstein, S. M. McClure, J. D. Cohen, *Trends Cogn. Sci.* **10**, 108 (2006).
27. R. Axelrod, D. Dion, *Science* **242**, 1385 (1988).
28. J. Bendor, R. M. Kramer, S. Stout, *J. Conflict Resolut.* **35**, 691 (1991).
29. D. Fudenberg, E. Maskin, *Am. Econ. Rev.* **80**, 274 (1990).
30. J. Wu, R. Axelrod, *J. Conflict Resolut.* **39**, 183 (1995).
31. M. A. Nowak, K. Sigmund, *Nature* **355**, 250 (1992).
32. C. Wedekind, M. Milinski, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 2686 (1996).
33. E. Fehr, C. F. Camerer, *Trends Cogn. Sci.* **11**, 419 (2007).
34. To explore whether lower levels of tit-for-tat reciprocity may have contributed to such failures of cooperation, linear regressions of behavior between partners were carried out. Consistent with previous work, reciprocity was found to be a strong predictor of subsequent increase or decreases in amount of money sent (table S2). Reciprocity is defined as the fractional change in money sent across rounds by one player in response to the fractional change in money sent across rounds by their partner. Thus, investor reciprocity on round j was calculated as $\Delta I_j - \Delta R_{j-1}$, where ΔI_j indicates the fractional change in investment from round $j-1$ to round j and ΔR_{j-1} indicates the fractional change in repayment from round $j-2$ to round $j-1$. Although deviation in perfect tit-for-tat behavior was a significant predictor of change in partner behavior across groups, the strength of this relation did not differ between pairs that included healthy trustees and pairs that included BPD trustees (table S2), which suggests sensitivity and responsivity to tit-for-tat behavior is intact among individuals with BPD.
35. B. King-Casas *et al.*, *Science* **308**, 78 (2005).
36. B. Seymour, T. Singer, R. Dolan, *Nat. Rev. Neurosci.* **8**, 300 (2007).
37. A. D. Craig, *Curr. Opin. Neurobiol.* **13**, 500 (2003).
38. A. D. Craig, *Annu. Rev. Neurosci.* **26**, 1 (2003).
39. P. R. Montague, T. Lohrenz, *Neuron* **56**, 14 (2007).
40. M. P. Paulus, *Science* **318**, 602 (2007).
41. C. Camerer, *Behavioral Game Theory: Experiments on Strategic Interaction* (Princeton Univ. Press, Princeton, NJ, 2003).
42. E. Fehr, K. M. Schmidt, *Q. J. Econ.* **114**, 817 (1999).
43. W. Güth, R. Schmittberger, B. Schwarze, *J. Econ. Behav. Organ.* **3**, 367 (1982).
44. G. S. Berns *et al.*, *Science* **312**, 754 (2006).
45. H. D. Critchley, C. J. Mathias, R. J. Dolan, *Neuron* **29**, 537 (2001).
46. S. A. Huettel, C. J. Stowe, E. M. Gordon, B. T. Warner, M. L. Platt, *Neuron* **49**, 765 (2006).
47. A. Simmons, S. C. Matthews, M. P. Paulus, M. B. Stein, *Neurosci. Lett.* **430**, 92 (2008).
48. M. P. Paulus, C. Rogalsky, A. Simmons, J. S. Feinstein, M. B. Stein, *Neuroimage* **19**, 1439 (2003).
49. K. Preuschoff, P. Bossaerts, S. R. Quartz, *Neuron* **51**, 381 (2006).
50. K. Preuschoff, P. Bossaerts, *Ann. N. Y. Acad. Sci.* **1104**, 135 (2007).
51. K. Preuschoff, S. Quartz, P. Bossaerts, *J. Neurosci.* **28**, 2745 (2008).
52. D. M. Dougherty, J. M. Bjork, H. C. G. Huckabee, F. G. Moeller, A. C. Swann, *Psychiatry Res.* **85**, 315 (1999).
53. K. M. Putnam, K. R. Silk, *Dev. Psychopathol.* **17**, 899 (2005).
54. S. Sieswerda, A. Arntz, M. Wolfis, *J. Behav. Ther. Exp. Psychiatry* **36**, 209 (2005).
55. J. B. Rotter, *J. Pers.* **35**, 651 (1967).
56. A. Todorov, L. T. Harris, S. T. Fiske, *Brain Res.* **1079**, 76 (2006).

57. This work was supported by the Child and Family Program at the Menninger Clinic, E. Bleiberg, Director (P.F., L.L., C.S), National Institute of Mental Health (NIMH) grant MH078485 (B.K.); NIMH grant MH52797 (P.R.M.); National Institute of Neurological Disorders and Stroke grant NS045790 (P.R.M.); and National Institute on Drug Abuse grant DA11723 (P.R.M). We thank P. Chiu for comments on this manuscript. We also thank R. Bhimani, C. Bracero, C. Ha, A. Jarman, O. Mosko, and technical staff of the Human Neuroimaging Laboratory at Baylor College of Medicine. We thank referring clinicians affiliated with the Menninger Department of Psychiatry at Baylor College of Medicine for assistance in recruitment, including E. Bleiberg, G. Gabbard, S. Twemlow, and E. Weinberg.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5890/806/DC1
Materials and Methods
Figs. S1 and S6
Tables S1 to 5
References

25 February 2008; accepted 30 June 2008
10.1126/science.1156902

The Crystal Structure of a Sodium Galactose Transporter Reveals Mechanistic Insights into Na^+/Sugar Symport

Salem Faham,[1] Akira Watanabe,[1] Gabriel Mercado Besserer,[1] Duilio Cascio,[2] Alexandre Specht,[3] Bruce A. Hirayama,[1] Ernest M. Wright,[1]* Jeff Abramson[1]*

Membrane transporters that use energy stored in sodium gradients to drive nutrients into cells constitute a major class of proteins. We report the crystal structure of a member of the solute sodium symporters (SSS), the *Vibrio parahaemolyticus* sodium/galactose symporter (vSGLT). The ~3.0 angstrom structure contains 14 transmembrane (TM) helices in an inward-facing conformation with a core structure of inverted repeats of 5 TM helices (TM2 to TM6 and TM7 to TM11). Galactose is bound in the center of the core, occluded from the outside solutions by hydrophobic residues. Surprisingly, the architecture of the core is similar to that of the leucine transporter (LeuT) from a different gene family. Modeling the outward-facing conformation based on the LeuT structure, in conjunction with biophysical data, provides insight into structural rearrangements for active transport.

A central question in biology is how energy is harnessed to do work. For the active accumulation of glucose into cells, Crane proposed that energy was obtained from the inward Na^+ gradient, that is, Na^+/glucose co-transport (symport) (*1*). This hypothesis has been extensively tested, refined, and expanded to include the active transport of solutes and ions in virtually all cell types (*2*). It is now established that cells maintain a low intracellular [Na^+] through the active pumping of Na^+ out of the cell. This inward Na^+ gradient, along with a negative membrane potential, drives the transport of substrates into cells. Despite this paradigm, the structural basis for Na^+ solute symport is unknown.

Solute sodium symporters (SSS) (TC# 2.A.21) are a large family of proteins that cotransport Na^+ with sugars, amino acids, inorganic ions, or vitamins (*3*). Members of this family are important in human physiology and disease where mutations in glucose and iodide symporters (SGLT1 and NIS) result in the congenital metabolic disorders glucose-galactose malabsorption (GGM) and iodide transport defect (ITD) (*4*, *5*). SGLT1 is the rationale for oral rehydration therapy, and SGLTs are currently being targeted in drug trials for type II diabetes.

The first member of the SSS family to be cloned was the intestinal Na^+/glucose symporter

[1]Department of Physiology, David Geffen School of Medicine, University of California, Los Angeles, CA 90095–1751, USA. [2]UCLA–Department of Energy Institute of Genomics and Proteomics, University of California, Los Angeles, CA 90095, USA. [3]Laboratoire de Chimie Bioorganique, Université Louis Pasteur, CNRS UMR 7175 LC01, Faculté de Pharmacie, 74 Route du Rhin, 67401 Illkirch, France.

*To whom correspondence should be addressed. E-mail: ewright@mednet.ucla.edu (E.M.W.); jabramson@mednet.ucla.edu (J.A.)

(SGLT1) (*6*), and since then more than 250 other members have been identified across all six kingdoms of life. The functions of the family members have been well characterized, including the mammalian glucose (SGLTs), the iodide (NIS), the *Vibrio parahaemolyticus* galactose/glucose (vSGLT), and the *Escherichia coli* proline (PutP) symporters (*3*, *7*, *8*). vSGLT has a sequence identity of 32% (60% similarity), 19% (58% similarity), and 18% (57% similarity) to SGLT1, NIS, and PutP, respectively (fig. S1). All share an alternating-access mechanism with tight coupling between Na^+ and solute transport (*9*–*11*). Despite functional studies supporting multiple states for symport, an absence of structural data on the protein family has precluded rigorous examination of this hypothesis.

To gain structural insight into the mechanistic details, we solved the structure of vSGLT in the presence of Na^+ and galactose. As predicted by experimental and in silico studies, vSGLT has 14 membrane-spanning [or transmembrane (TM)] helices with extracellular amino and carboxy termini (*12*, *13*). The structural core is formed from inverted repeats of 5 TM helices (TM2 to TM6 and TM7 to TM11), placing it in an inward-facing conformation. Galactose is bound in the center of the core, occluded from the outside solutions by hydrophobic residues. We use structural and functional data, coupled with modeling, to predict the conformational changes as the transporter alternates between the outward- and inward-facing conformations, thereby providing a structural explanation for symporter dynamics.

Structure determination. Optimized crystals of SGLT from *V. parahaemolyticus* displayed anisotropic diffraction, with Bragg spacing of 2.7 Å and 3.7 Å along the best and worst directions, respectively. Data were collected for three different crystal forms: P1, $P2_1$, and $P2_12_12_1$. Initial phases were obtained in the $P2_1$ crystal form, permitting the building of a partial model that we used for molecular replacement into the P1 and $P2_12_12_1$ crystal forms. Multicrystal averaging and solvent flattening substantially improved experimental phases, and Se-methionine crystals were used for sequence assignment. The model was refined from merged data to a resolution of 2.7 Å with an R-free of 28.7% and an R-work of 27.0% (table S1). Of the 548 residues of vSGLT, 512 were built; the remaining residues are in disordered loop regions (*14*).

Transporter architecture. The protein assembles as a tightly packed parallel dimer, with an extensive contact surface area of 1488 $Å^2$ occurring between TM helices within the lipid bilayer (fig. S2). The functional relevance of the dimer is unclear, as freeze-fracture electron microscopic studies indicate that SGLTs function as monomers (*13*, *15*). The overall architecture has a maximal height and diameter of ~75 Å and ~55 Å (Fig. 1). There are 14 transmembrane helical segments (TM1 to TM14) with both N and C termini exposed to the periplasm. A single galactose molecule is bound in the center of the protomer. A structural characteristic, not anticipated from amino acid sequence, is the inverted topology repeat motif; TM2 to TM6 and TM7 to TM11 are related by a ~153° rotation parallel to the membrane plane and are structurally related (Fig. 1A and fig. S4, A and B). These 10 TM helices likely represent the core structure in the SSS family, with additional helices located on the N and C termini. This inverted topology repeat appears to be a common motif of membrane transport proteins (*16*–*20*).

The structure is composed of a central group of seven helices (TM2, TM3, TM4, TM7, TM8, TM9, and TM11) that contribute side-chain interactions for ligand selectivity, along with seven supporting helices that stabilize these central helices (Fig. 1 and fig. S3C). The positioning of the central helices in complex with the substrate forms a large cavity that extends from just below the galactose-binding site to the intracellular space, representing the inward-facing conformation (Fig. 4B). A striking feature is the two discontinuous TM helices, TM2 and the symmetrically related TM7, in the center of the protomer. In TM2 there is a break in the hydrogen-bonding pattern around residues I65, S66, and A67, dividing it into roughly equivalent intracellular (TM2I) and extracellular (TM2E) components. In TM7 the helix is disrupted at residues F266, N267, Q268, and Y269, dividing it into a shorter intracellular (TM7I) and a larger extracellular (TM7E) segment (Figs. 1 and 2A). This structural feature may have particular functional importance, as discontinuous helices have been implicated in transport mechanisms of several cotransporters (*18*, *20*, *21*).

There are notable structural features in the solvent-accessible regions. On the intracellular side, helix IL3—situated between TM3 and TM4 (Fig. 1)—is located on the outer edge of the cavity leading to the substrate-binding site. On the extracellular face between TM8 and TM9, there are two extracellular helices, EL8Ha and EL8Hb. There may be mechanistic implications with EL8Hb as it straddles the membrane plane and forms extensive contacts with the central helices TM2 and

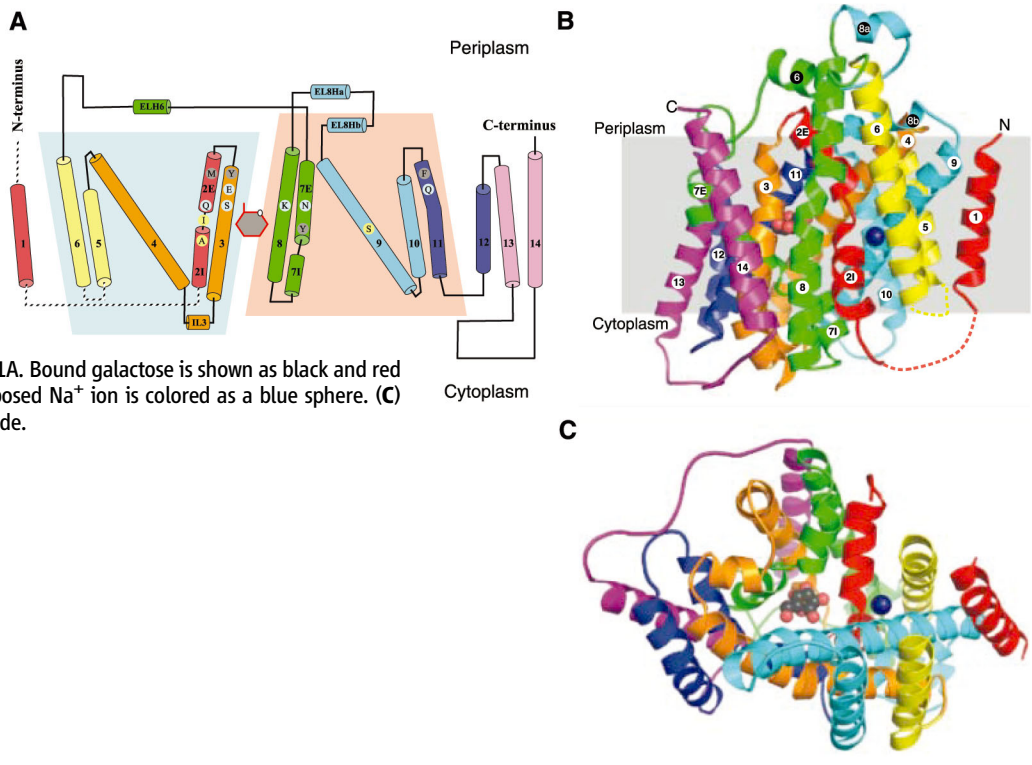


Fig. 1. Structure of vSGLT. (**A**) Topology. The structure is colored as a rainbow from the N terminus (red) to the C terminus (purple). The blue and red trapeziums represent the inverted topology of TM2 to TM6 and TM7 to TM11. The gray hexagon with red trim represents the galactose. Residues involved in sugar recognition, gate residues, and a proposed Na^+ site are shown in cyan, gray, and yellow circles. (**B**) Structure viewed in the membrane plane. The coloring scheme and numbering of α helices is the same as in Fig. 1A. Bound galactose is shown as black and red spheres for the C and O atoms. The proposed Na^+ ion is colored as a blue sphere. (**C**) Structure viewed from the intracellular side.

TM4. There is an additional extracellular helix (ELH6) between TM6 and TM7. This loop connects the two inverted repeats and harbors a glycosylation site in the eukaryotic SGLTs (*22*).

The 10 core helices (TM2 to TM11) form a structure that displays a similar topology to the core structure of LeuT (TM1 to TM10), a bacterial member of the neurotransmitter sodium symporters (NSS) family (*18*) (fig. S4C). This structural homology was surprising, because the SSS and NSS families have no significant similarity at the sequence level. Sequence-based programs were unable to find the alignment between LeuT and vSGLT, and a structure-based alignment displayed only an 11.5% identity. These findings support classification of proteins using criteria such as topological arrangement, molecular function, and unique structural features involved in mechanism, rather than solely on the basis of primary sequence (*23*).

Galactose-binding site. Galactose analogs modified with electron-dense atoms were used to identify the precise position and orientation of the sugar molecule. Galactose is bound about halfway across the membrane bilayer by specific side-chain interactions from the central helices TM2E, TM3, TM7E, TM8, and TM11 (Fig. 2). The galactose-binding site is sandwiched between hydrophobic residues that form intracellular and extracellular gates. On the intracellular side, Y263 from the discontinuous helix TM7E stacks with the pyranose ring, a common feature seen in all sugar-binding structures to date (*24*, *25*). This primary interaction, along with the flanking residues Y262 and W264, establishes a gate that prevents exit of the sugar to the large hydrophilic cavity contiguous with the intracellular compartment. The extracellular gate is formed by a triad of hydrophobic residues (M73, Y87, and F424) (Fig. 2B). Directly above the hydrophobic residues, there are substantial interactions between TM11, TM3, TM2E, and TM7E with the loops from TM2-TM3, TM8-TM9, and TM10-TM11. Taken together, these interactions comprise considerable protein mass that extends from the substrate-binding site to the extracellular surface, thereby further blocking access of the sugar-binding region from the extracellular milieu. In the members of the SSS family that bind sugars, the residues forming the hydrophobic gates are highly conserved (fig. S1).

All OH-groups of the galactose ring are coordinated through H-bonds (Fig. 2C). A possible H-bond between the carbonyl oxygen of Y87 and the C4-OH of galactose is also observed. Because Y87 is part of the extracellular hydrophobic gate, this additional H-bond may have mechanistic implications for galactose binding and stabilization in the central domain. K294 forms a H-bond with the C2-OH in a coordination similar to that observed in the glucose/galactose binding protein (PDB 1GCA) and lactose permease (PDB 1PV7); indeed, H-bonding with a positively charged side chain is a common feature of sugar-binding proteins (*24*–*26*). Overall, vSGLT supports the sugar by extensive side-chain interactions from the central domain and is held in an occluded conformation by intracellular and extracellular hydrophobic gates.

To verify the structural data and understand the functional importance of substrate binding, most of the residues forming the sugar-binding site were mutated to alanine, and galactose transport was measured in proteoliposomes (Fig. 2D). The Q69A, E88A, N260A, K294A, and Q428A mutations did not show Na^+-dependent galactose transport. In contrast, the S91A mutant behaved normally, which suggests that it is not essential for protein function (S91 is not conserved across different species). Mutants of the corresponding sugar-binding site residues in SGLT1 (K321A, Q457R, and Q457C) (fig. S1A) provide strong support for the argument that these residues also play an important role in the mammalian transporter (*27*, *28*). The mutant Q457R causes severe glucose-galactose malabsorption in human subjects, a potentially fatal disease in newborn infants (*4*). This indicates a conserved role for these sugar-coordinating residues in protein function across isoforms and species.

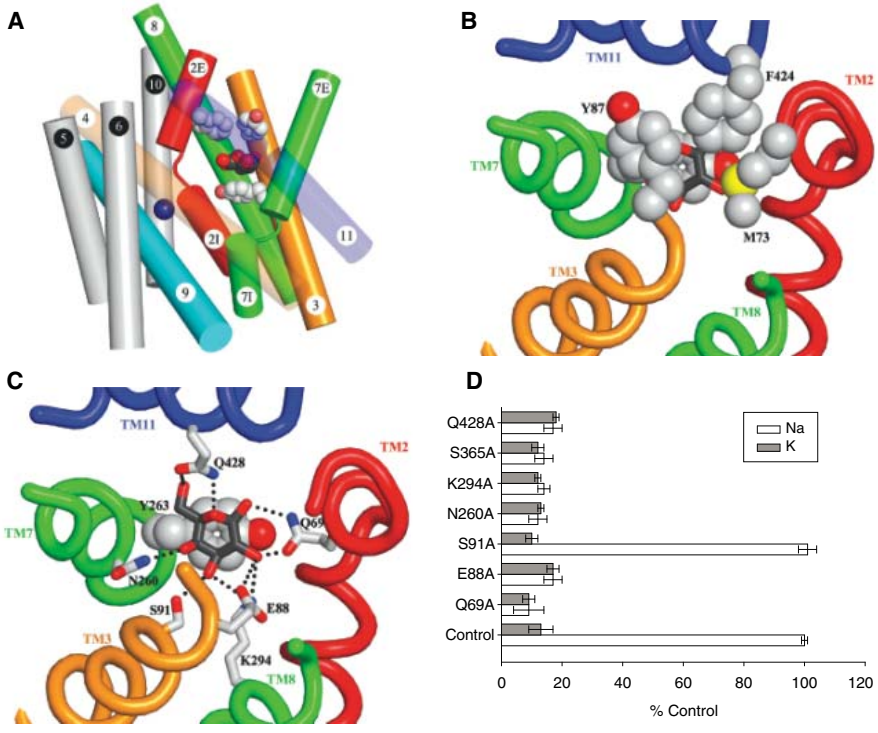


Fig. 2. Galactose-binding site. (**A**) Overview of the galactose and proposed Na^+-binding site viewed in the membrane plane, maintaining the same color scheme as in Fig. 1. (**B**) Hydrophobic gate residues (viewed from the extracellular side). The intracellular (Y263) and extracellular (M73, Y87, and F424) gates are shown as spheres, and the galactose is shown as sticks. (**C**) The galactose-binding site [same view as in (B)], with the extracellular hydrophobic gate residues removed to view the galactose-binding site. Residues in the galactose-binding site are displayed as sticks colored by atom type. Hydrogen bonds are depicted as black dashed lines. (**D**) D-galactose transport by vSGLT mutants in proteoliposomes. The results are expressed as % Control of that measured for p3C423 in 100 mM NaCl (~1.2 nmoles min^{-1} mg^{-1} protein). SEM is displayed for each experiment.

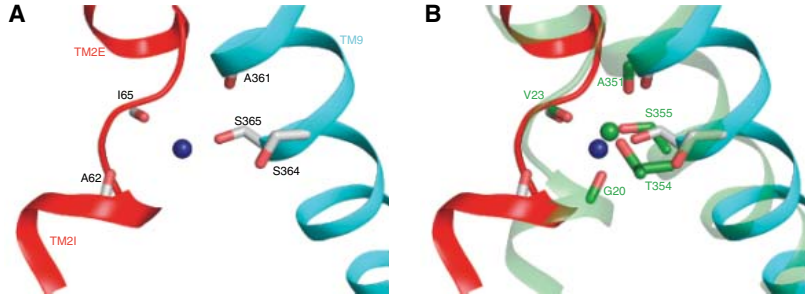


Fig. 3. The proposed Na^+-binding site. (**A**) Residues in the Na^+-binding site are displayed as sticks colored by atom type, and the helices are colored as in Fig. 1. (**B**) Superposition of the proposed Na^+ site of vSGLT onto the second Na^+ site from the LeuT structure (green). Alignment was performed using only helices 2 and 9 from vSGLT and the corresponding helices from LeuT.

Sodium-binding site. Although the Na^+ electrochemical gradient is the driving force for symport, Na^+/substrate stoichiometry varies amongst SSS members; for SGLT1 and NIS, the stoichiometry is 2:1 (*10*, *11*), whereas for vSGLT and PutP, it is 1:1 (*8*, *13*). Nevertheless, it is expected that the mechanisms that couple transport of Na^+ and substrate will have commonality within the SSS family.

As in other transporters, precise localization of the Na^+ binding site(s) in vSGLT is not easily identified because of the similarity in density between Na^+ and water. Cocrystallization was attempted with electron-dense ions (Rb, Cs, and Tl), but none were observed in the resulting structure. Nonetheless, we were able to identify a plausible Na^+-binding site based on a comparison with the LeuT structure (*18*), conservation of sequence amongst SSS proteins, and a mutational analysis. Structural alignment with LeuT revealed a possible Na^+-binding site at the intersection of TM2 and TM9, ~10 Å away from the substrate-binding site (Fig. 3). In this region, the carbonyl oxygens from A62 (TM2E), I65 (discontinuous segment of TM2), and A361 (TM9) would be appropriately positioned to coordinate the Na^+. An interaction is also possible with the hydroxyl oxygen of S365 of TM9, a residue conserved throughout the SSS family (fig. S1). This site has a similar proximity, as well as coordination properties, to the second Na^+ site proposed for LeuT (*18*): Superimposition of the two structures reveals localization of these Na^+ sites in the two protein structures within ~2 Å (Fig. 3B). Functional studies provide supporting evidence for Na^+ binding in this region. In PutP, a hydroxyl moiety at position T341 is essential for Na^+ binding (*29*), and studies on NIS have shown that T354 plays a key role in Na^+-dependent transport of iodide (*30*). Both of these residues are at equivalent position to S365 on TM9 in vSGLT. Mutation S365A in vSGLT completely abrogated Na^+-dependent transport (Fig. 2D). Unlike LeuT, but consistent with biochemical data, we did not detect a second Na^+ binding site in vSGLT.

The inward-facing conformation. vSGLT resides in an inward-facing conformation, with the galactose-binding site inaccessible to the extracellular and intracellular compartments (Fig. 2). The intracellular exit pathway appears as a large hydrophilic cavity blocked by the intracellular gate residue (Y263 on TM7E) (Fig. 2C). The cavity is formed from the intracellular portions of helices TM2I, TM3, TM4, TM7I, TM9, and TM11 (Fig. 4). This finding is consistent with accessibility studies on PutP (*29*). Cysteine-scanning mutagenesis on TM2 and TM9 of PutP reveals a hydrophilic face accessible to bulky sulfhydryl reagents (fluorescein-5-maleimide), indicating an aqueous pathway between the site of substrate-binding and the cytoplasm (*29*, *31*). For NIS, five hydroxyl-containing residues (T351, S353, T354, S356, and T357) and N360 located along one face of TM9 play a key role in Na^+/iodide cotransport (*30*). Together, these data suggest that the inward-facing conformation of vSGLT hosts a large cavity exposed to the cytoplasm that requires simple displacement of an intracellular gating residue (Y263) for release of galactose.

Alternating access. According to kinetic models for Na^+/solute symport (*8*–*10*), the proteins alternate between outward- and inward-facing conformations. Thus, for galactose to enter the vSGLT binding site from the extracellular side, it must adopt an outward-facing conformation. Clues about this conformation arise from the structure of LeuT (*18*), which is in an outward-facing conformation. Given the structural similarity between vSGLT and LeuT, we mapped the core domain of vSGLT onto that of LeuT using the program MODELLER (*32*). Comparison of the two conformations (Fig. 4, A and B) in conjunction with biochemical data indicates a series of specific helical rearrangements and a plausible pathway for Na^+ and galactose entry. Specifically, modeling reveals an external pathway to the substrate-binding site formed by TM2E, TM3, TM7E, TM11, and helix EL8b in the extracellular loop. External Na^+ binds first (*33*), presumably at the site identified in Fig. 3, and we postulate that this facilitates molecular rearrangements in TM2 to form the substrate-binding site. A possible link between the Na^+ and substrate-binding sites is residue N64, which is in hydrogen-bonding distance to the C2-OH of galactose. Galactose binding will induce the formation of the extracellular gate (Y87, F424, and M73), closing the cavity through bends in TM3 and TM11. These structural rearrangements are facilitated by conserved glycine and proline residues (TM3 G99 and P104; TM11 P436 and G437). Consistent with this model, the corresponding helices of LeuT have conserved glycine residues in the same regions. Further evidence for these structural rearrangements is corroborated by site-specific fluorescence experiments on vSGLT, hSGLT, and PutP (*8*, *33*–*35*).

Larger conformational change occurs in the intracellular portions of TM segments TM3 and TM4, TM7 and TM8, and TM9 to TM11 to expand the intracellular cavity. This places the protein in the state represented by our structural data: inward-facing but blocked by the internal hydrophobic gate (Y263). Although the residency time in the occluded state is unknown, the protein must next shift conformation to release Na^+ and then galactose by displacing the intracellular gate (i.e., displacement of Y263). Supporting this mechanism are studies on hSGLT that document a conformational change in the protein after sugar binding (*9*), which may be indicative of either occlusion of the external cavity and/or exposure of the internal cavity.

The structure of vSGLT in an inward-facing conformation, along with the structure of the functional and structural homolog LeuT, provides a model of the conformational changes underlying the alternating access mechanism. These structural and functional similarities indicate a common transport mechanism between members of the SSS and NSS gene families. To expand on this transport mechanism, further structural and functional analyses are required. Finally, given the functional homology between the bacterial and human SGLTs, the vSGLT structure should provide a critical tool in the rational design of SGLT drugs for the treatment of diabetes.

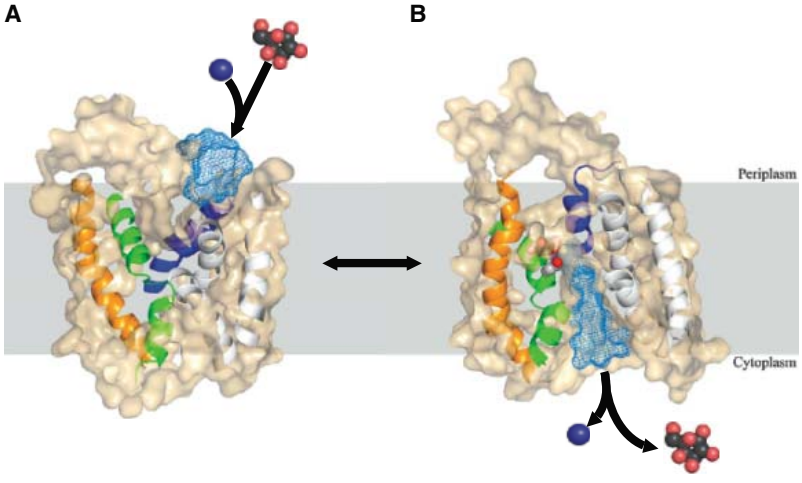


Fig. 4. Alternating accessibility. (**A**) Slice through surface of the outward-facing model viewed from the membrane plane showing the extracellular cavity (blue mesh). (**B**) Slice through surface of the inward-facing structure of vSGLT in the membrane plane showing the intracellular cavity (blue mesh). Helices showing structural rearrangement are colored orange, green, and blue for helices TM3, TM7, and TM11; helices with little movement are colored white. The surface is shown in beige. Galactose is shown as black and red spheres for the C and O atoms, respectively. Y263 and the Na^+ ion are colored as gray and blue spheres, respectively.

References and Notes

1. R. Crane, D. Miller, I. Bihler, *Membrane Transport and Metabolism*, A. Kleinzeller, A. Kotyk, Eds. (Academic Press, London, 1961), pp. 439–449.
2. S. G. Schultz, P. F. Curran, *Physiol. Rev.* **50**, 637 (1970).
3. E. M. Wright, D. D. Loo, B. A. Hirayama, E. Turk, *Physiology (Bethesda)* **19**, 370 (2004).

4. E. M. Wright, B. A. Hirayama, D. F. Loo, *J. Intern. Med.* **261**, 32 (2007).
5. M. D. Reed-Tsur, A. De la Vieja, C. S. Ginter, N. Carrasco, *Endocrinology* **149**, 3077 (2008).
6. M. A. Hediger, M. J. Coady, T. S. Ikeda, E. M. Wright, *Nature* **330**, 379 (1987).
7. O. Dohan *et al.*, *Endocr. Rev.* **24**, 48 (2003).
8. H. Jung, *FEBS Lett.* **529**, 73 (2002).
9. D. D. Loo, B. A. Hirayama, M. H. Karakossian, A. K. Meinild, E. M. Wright, *J. Gen. Physiol.* **128**, 701 (2006).
10. S. Eskandari *et al.*, *J. Biol. Chem.* **272**, 27230 (1997).
11. B. Mackenzie, D. D. Loo, E. M. Wright, *J. Membr. Biol.* **162**, 101 (1998).
12. E. Turk, E. M. Wright, *J. Membr. Biol.* **159**, 1 (1997).
13. E. Turk *et al.*, *J. Biol. Chem.* **275**, 25711 (2000).
14. Materials and methods are available as supporting material on *Science* Online.
15. S. Eskandari, E. M. Wright, M. Kreman, D. M. Starace, G. A. Zampighi, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11235 (1998).
16. D. Fu *et al.*, *Science* **290**, 481 (2000).
17. R. Dutzler, E. B. Campbell, M. Cadene, B. T. Chait, R. MacKinnon, *Nature* **415**, 287 (2002).
18. A. Yamashita, S. K. Singh, T. Kawate, Y. Jin, E. Gouaux, *Nature* **437**, 215 (2005).
19. J. U. Bowie, *Nat. Struct. Mol. Biol.* **13**, 94 (2006).
20. C. Hunte *et al.*, *Nature* **435**, 1197 (2005).
21. D. Yernool, O. Boudker, Y. Jin, E. Gouaux, *Nature* **431**, 811 (2004).
22. M. A. Hediger, J. Mendlein, H. S. Lee, E. M. Wright, *Biochim. Biophys. Acta* **1064**, 360 (1991).
23. L. N. Kinch, N. V. Grishin, *Curr. Opin. Struct. Biol.* **12**, 400 (2002).
24. M. S. Sujatha, P. V. Balaji, *Proteins* **55**, 44 (2004).
25. J. Abramson *et al.*, *Science* **301**, 610 (2003).
26. J. Y. Zou, M. M. Flocco, S. L. Mowbray, *J. Mol. Biol.* **233**, 739 (1993).
27. D. D. Loo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7789 (1998).
28. M. Panayotova-Heiermann, D. D. Loo, J. T. Lam, E. M. Wright, *Biochemistry* **37**, 10522 (1998).
29. D. Hilger, M. Bohm, A. Hackmann, H. Jung, *J. Biol. Chem.* **283**, 4921 (2008).
30. A. De la Vieja, M. D. Reed, C. S. Ginter, N. Carrasco, *J. Biol. Chem.* **282**, 25290 (2007).
31. T. Pirch, S. Landmeier, H. Jung, *J. Biol. Chem.* **278**, 42942 (2003).
32. N. Eswar *et al.*, *Curr. Protoc. Bioinformatics Suppl.* **15**, 5.6.1. (2006).
33. M. Veenstra, S. Lanza, B. A. Hirayama, E. Turk, E. M. Wright, *Biochemistry* **43**, 3620 (2004).
34. E. Turk, O. K. Gasymov, S. Lanza, J. Horwitz, E. M. Wright, *Biochemistry* **45**, 1470 (2006).
35. B. A. Hirayama *et al.*, *Biochemistry* **46**, 13391 (2007).

36. The authors thank E. Turk for contributions to the early stages of the project; J. Whitelegge for assistance with the electrospray mass spectroscopy; and R. Kaback, K. Philipson, J. Bowie, T. Vondriska, V. Chaptal, and D. Loo for critical comments on the manuscript. M. Hahn and D. Kwon helped in sample preparation, and the Protein Expression Core made large-scale protein production possible. The Advance Light Source, Advance Photon Source, and Swiss Light Source synchrotron were all used. The authors would like to emphasize the efforts of P. Zwart and the beamline staff at 5.0.2. The coordinates are deposited at the Protein DataBank, accession code 3DH4. Supported by NIH grants DK19567 (E.M.W.), DK44602 (E.M.W.), and NIH GM07844 (J.A.), HFSP RGY0069 (J.A.), and AHA 0630258N (J.A.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/1160406/DC1
Materials and Methods
Figs. S1 to S4
Table S1
References

13 May 2008; accepted 18 June 2008
Published online 3 July 2008;
10.1126/science.1160406
Include this information when citing this paper.

REPORTS

Gas Disks to Gas Giants: Simulating the Birth of Planetary Systems

Edward W. Thommes,[1,2]* Soko Matsumura,[2] Frederic A. Rasio[2]

The ensemble of now more than 250 discovered planetary systems displays a wide range of masses, orbits and, in multiple systems, dynamical interactions. These represent the end point of a complex sequence of events, wherein an entire protostellar disk converts itself into a small number of planetary bodies. Here, we present self-consistent numerical simulations of this process, which produce results in agreement with some of the key trends observed in the properties of the exoplanets. Analogs to our own solar system do not appear to be common, originating from disks near the boundary between barren and (giant) planet-forming.

Gas giant planets form within 1 to 10 million years (My), during the time that the parent star possesses a gas disk (*1*). Two- and three-dimensional hydrodynamic simulations of planets embedded in protostellar disks are, due to their computational cost, limited to following a maximum of 10^3 to 10^4 orbits; even for a planet as far out as Jupiter (orbital period 11 years), this amounts to at most a tenth of a protostellar disk's total lifetime. For this reason, longer time-scale simulations usually model the time after the protostellar gas disk has dissipated, allowing the problem to be tackled with pure gravitational *N*-body methods. Recent work in this area has been successful at reproducing the observed exoplanet eccentricity distribution, the key requirement being simply that planets must begin close enough to each other to render systems dynamically active from the outset (*2*, *3*). Although these results provide support for planet-planet interaction as the main agent behind the eccentricities, they do not address the question of how, or even if, planet formation produces the requisite crowded systems. Furthermore, many of the observed cases appear to have kept a memory of their early dynamical evolution: About a quarter of discovered systems contain planet pairs locked into dynamical mean-motion resonances (*4*), likely the product of early differential migration (*5*) by planet-disk interaction (*6*–*8*). Thus, to achieve a more complete understanding of the planet formation process, it is essential to bridge the disk and postdisk era. Most models of this regime have been either semianalytic treatments, which do not consider planet-planet interactions (*9*–*12*), or *N*-body simulations with parameterizations of disk effects (*13*, *14*). We performed simulations using a hybrid numerical scheme (*15*), which combines the N-body code SyMBA (Symplectic Massive Body Algorithm) (*16*) with a one-dimensional disk model that self-consistently interacts with, and accretes onto, the embedded planets. We began our simulations with the appearance of the first small protoplanets (10^{-3} M_E, where M_E is Earth's mass) in a gas disk and ran them for up to half a billion years (Gy). Gas giants are assumed to form by core accretion [see supporting online material (SOM) for a more detailed description of the code].

The fundamental question we sought to address is how the properties of a mature planetary system map to those of its birth disk. To this end, we performed simulations covering a range of disk parameters. Fig. 1 depicts one representative example, which illustrates some of the key effects at work in the planet formation process (see Movie S1 for an animated version). In this simulation, protoplanet growth initially proceeded in an orderly manner; the dissipational effect of the gas disk dominated over the bodies' mutual perturbations, so that all orbits were kept nearly circular. Simultaneously, interaction with the disk also caused protoplanets to undergo orbital decay, referred to as "type I migration" (*8*). This effect will in general deposit some protoplanets at the inner disk edge (or onto the central star) before they have a chance to become gas giants (*10*), but because migration speed scales with migrator mass and gas disk density, those protoplanets that grow slowly enough relative to the disk's depletion time avoid this fate (*17*, *12*). All protoplanets in the simulation (Fig. 1) fell into the latter category (see SOM for counterexamples), although the orbit of the innermost one, originating at 2 astronomical units (AU), shrank to 1/5 its original radius by 1 My. Around this time, the first gas giant formed, scattering many of the neighboring protoplanets to high eccentricities as its mass rapidly grew. Its perturbation on the surrounding gas also became strong, and a deep

[1]University of Guelph, Guelph, ON N1G 2W1, Canada. [2]Northwestern University, Evanston, IL 60208, USA.

*To whom correspondence should be addressed; E-mail: thommes@northwestern.edu

annular gap opened around its orbit. With gas unable to readily cross its orbit, the planet was locked into the disk and carried along by the latter's accretion flow, in what is termed "type II migration" (*8*). At 1.2 My, the gap became a hole as the inner disk material drained away faster than the planet migrated; however, the hole rapidly closed in again, depositing the planet at the inner disk edge at 1.24 My. By this time, two more gas giants began to form. Each in turn opened a hole in the disk, in the process cutting off its inner neighbor from the gas, so that ultimately only the outermost planet was directly interacting with, and accreting from, the disk. The outer planet migrated toward the middle one, and at 1.66 My, the two become locked into a 3:1 mean-motion resonance (*18*) with each other, mutually increasing their eccentricities as they moved inward together. Migration and further planet accretion stalled as the disk density became low; the disk disappeared altogether a little after 4 My, leaving behind a system consisting of three gas giants—the outer two still in resonance—plus an outer Neptune-class planet. This system remained stable for the rest of the simulation, which ended after 0.5 Gy.

We performed a set of 100 simulations over a range of parameters. The properties of protoplanetary disks are only weakly constrained; two key parameters are initial gas disk mass M_{disk}, and disk viscosity ν, which determines how rapidly accretion onto the central star removes the bulk of the gas [photoevaporation may remove the last of the gas (*19*)]. Observations combined with modeling (*20*) suggest that $\sim 10^{-2}$ $M_{Sun} < M_{disk} < 10^{-1}$ M_{Sun} and $10^{-3} < \alpha < 10^{-2}$, using the common parameterization (*21*) $\nu \equiv \alpha c_s H$, with c_s the gas sound speed and H the disk scale height. We used this range of parameters for our simulations; the resulting planetary systems (Fig. 2) were each the product of a complex interplay between planet-disk and planet-planet effects; thus, individual outcomes were highly stochastic. Nevertheless, clear trends with M_{disk} and α are visible: In one extreme, low disk mass combined with high viscosity resulted in systems that produced no gas giants at all. In the other extreme, high disk mass combined with low viscosity results in the production of numerous gas giants; most underwent considerable inward migration, and many acquired large eccentricities.

We can understand these results in terms of two fundamental time scales in a planet-forming disk: One is the gas disk depletion time τ_{disk}; the other is the time to form the first gas giant, τ_{giant}. As shown in Fig. 3, we expect an initial burst of planet formation that spreads out from a particular radius, typically comparable to the Jupiter-Saturn region of our own solar system (5 to 10 AU). As time passes, the delay between the birth of successive giants becomes longer and longer. Thus, the crucial factor determining how a given system's formation history will play out is the time during the gas disk's lifetime that this burst occurs. In cases with $\tau_{giant} > \tau_{disk}$ (Fig. 3, lower right), the gas is removed before any gas giant has a chance to form, leaving behind systems consisting solely of rocky-icy bodies. In cases with $\tau_{giant} < \tau_{disk}$ (upper-left region of Fig. 2), planets are born into a substantial gas disk, and such systems generally produced a number of gas giants that migrated inward a considerable distance. Planet-bearing systems can be further classified into "planet-dominated" ones, wherein the planets clear the disk from the inside out, and "disk-dominated" cases, in which the disk clears the planets, their ultimate fate depending on what happens at the star-disk interface (not resolved in our simulations). The system depicted in Fig. 1 represents an intermediate case, with more examples shown in the SOM. Whether planets accumulate at the original inner disk edge (*22*–*24*) or in an inner hole, migration tends to produce crowded systems, leading to the excitation of eccentricities by planet-planet scattering and resonances (*2*, *3*, *13*, *14*). We do not know where the true distribution of disk properties falls, but because 6 to 7% of Sun-like stars are observed

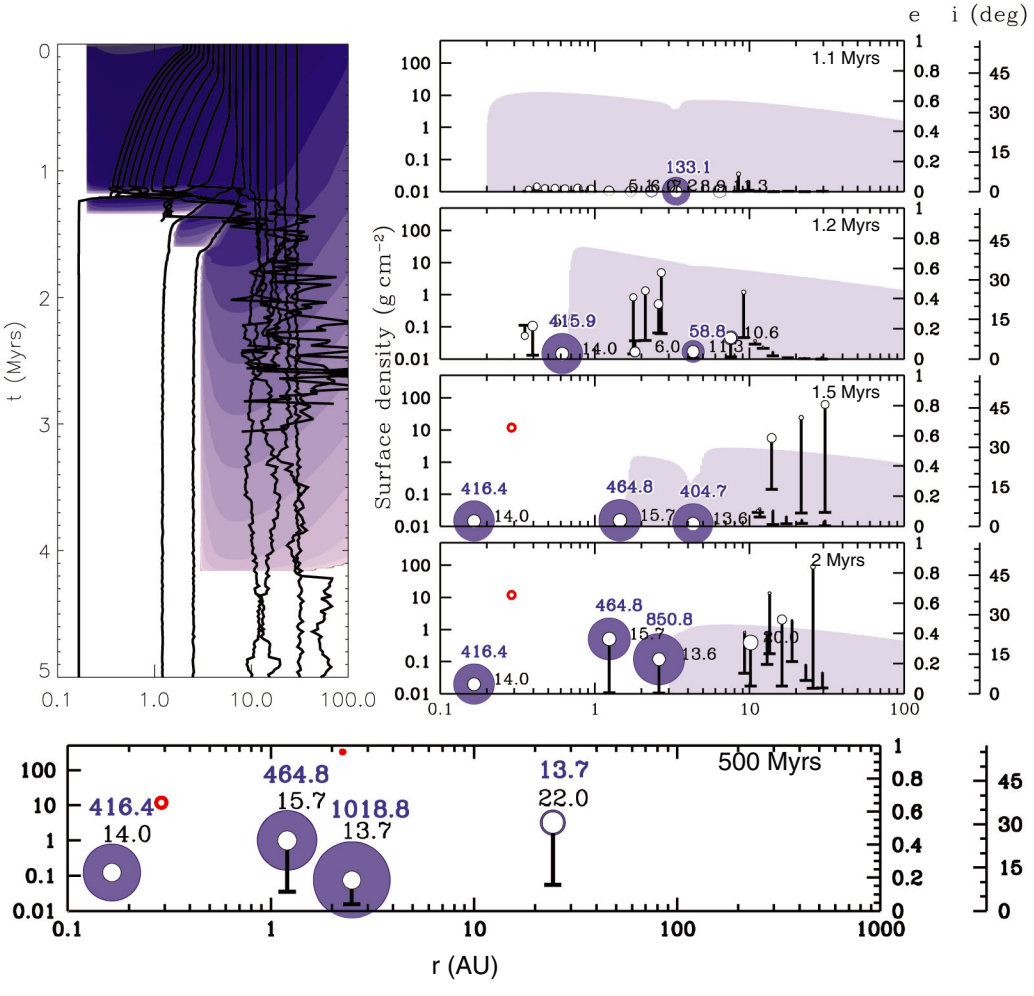


Fig. 1. Example of a planet formation simulation. A solar-metallicity disk with initial mass $M_{disk} = 0.088$ M_{Sun} and viscosity parameter $\alpha = 7.5 \times 10^{-3}$, containing protoplanets of initial mass 10^{-3} M_E between 2 and 30 AU, evolves for 0.5 Gy. (Top left) Planet semimajor axis over time, with the azimuthally averaged gas disk surface density overlaid as a contour plot. (Right, top four panels) Snapshots of the system at different intermediate times, showing planet eccentricity (right scale) and inclination (far-right scale, indicated by a horizontal tickmark connected to each planet by a vertical line) versus semimajor axis. Planet cores (black empty circles), together with whatever gas envelopes they have accreted (dark blue), are labeled with their mass in M_E (black and dark blue, respectively). Planets crossing the inner simulation boundary at 0.1 AU are removed, and their final mass and orbital elements shown (empty red circles). The azimuthally averaged disk surface density is also shown (light blue, left scale). One planet ends up at the inner disk edge [likely < 0.1 AU if due to the star's magnetospheric cavity (*22*); however, we set it at 0.2 AU for computational reasons]. (Bottom) The state of the system at 0.5 Gy. Orbital elements are averages over the last 1 My.

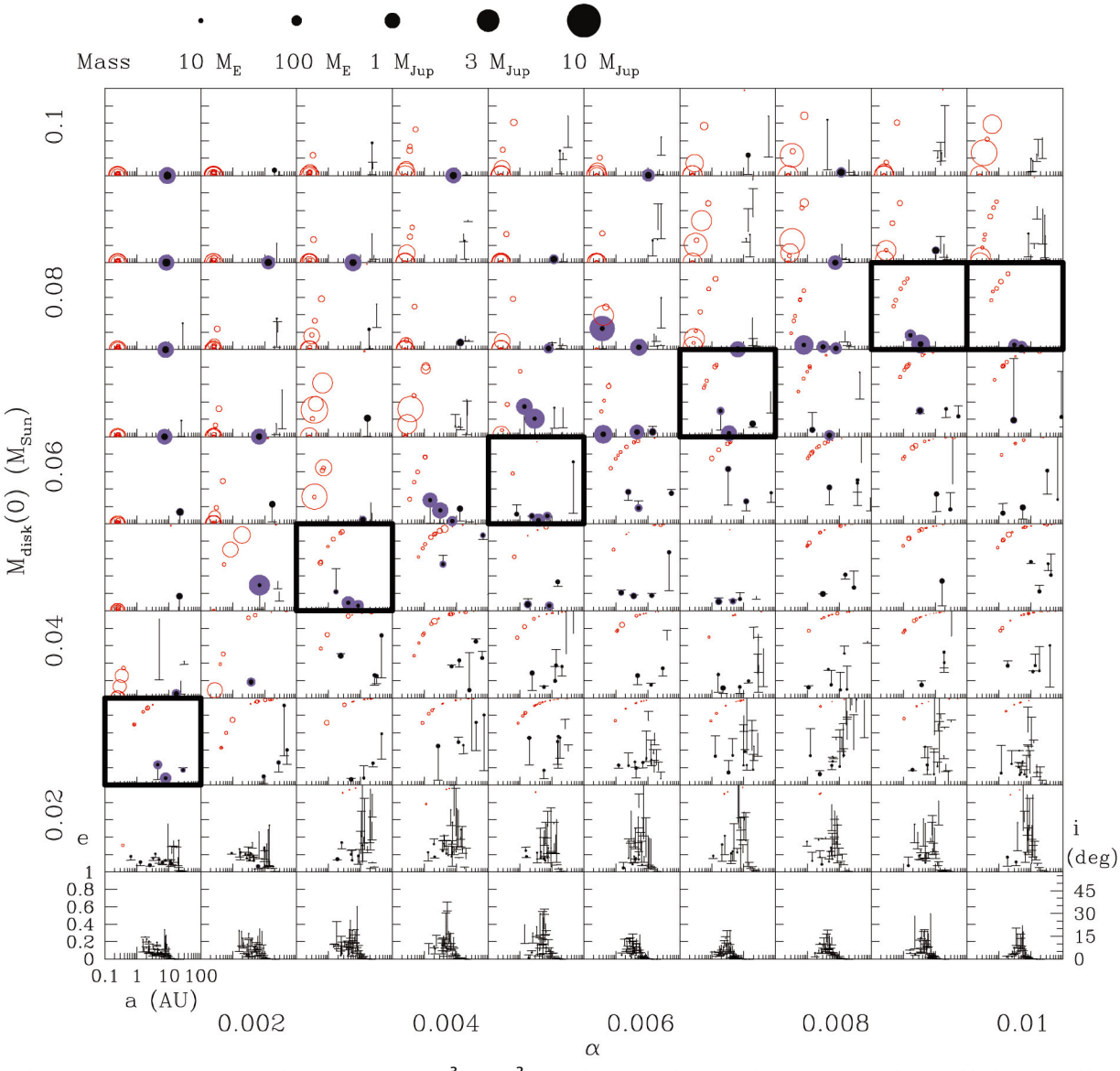


Fig. 2. Final outcomes of a set of 100 simulations, spanning 10^{-3} to 10^{-2} in viscosity parameter α, and 0.01 M_{Sun} to 0.1 M_{Sun} in initial disk mass. Simulations are ended after 0.5 Gy have elapsed, or if they fail to produce any gas giants within the gas disk's lifetime. Planet semimajor axes, eccentricities and inclinations (denoted as in Fig. 1) are plotted, as well as the relative solids and gas content of each planet [solid core, black; gaseous envelope, dark blue; size $\propto$ mass$^{1/3}$ (see mass scale at top)]. To keep computational cost reasonable, simulations have an inner boundary at 0.25 AU, beyond the initial inner edge of the gas disk; any body that crosses the boundary is removed, and a red circle is plotted showing its orbital elements and mass at the time of removal. Toward high M_{disk} and low α, planets form early and often during the gas disk's lifetime, most migrate extensively, and many acquire high eccentricities in the process (see Fig. 1 and SOM). Toward low M_{disk} and high α, planet formation is too slow to produce any gas giants during the disk lifetime. Between these two extremes is a relatively narrow boundary region (thick borders) in which gas giants migrate little and remain at low eccentricity, thus producing some outcomes more similar to the solar system. No gas giants at all form in disks of 0.02 M_{Sun} or less; for comparison, this is the approximate lower limit on the solar system's birth disk, called the "minimum mass solar nebula" (*37*).

to harbor giant planets (*25*), at least that fraction of disks must fall into the giant planet-bearing upper-left region of Fig. 2. Despite the speed of the hybrid code, it is as yet not computationally feasible to perform simulations of a sufficient fidelity and number to allow a detailed comparison to the statistical properties of the discovered exoplanets. Instead, we performed a smaller set of lengthier simulations, focusing principally on a part of the planet-bearing region. These produced qualitative agreement with some of the key features of the exoplanet ensemble, namely the correlation of host-star metallicity with planet occurrence (*26*, *27*), the distribution of planet masses, planet orbital periods, and the mass-eccentricity distribution (*28*), as well as insights into how these features arose (see SOM for details).

Our results also suggest how the solar system fits into the picture. In systems with $\tau_{giant} \sim \tau_{disk}$, gas giants do form but undergo only modest migration and eccentricity growth; thus, it is here where we would most naturally expect to find a solar system–like outcome. Figure 2 shows that these cases occupy a relatively narrow region within the parameter space, roughly a diagonal line extending from ($\alpha = 1 \times 10^{-3}$, $M_{disk} = 0.03$ M_{Sun}) to ($\alpha = 10^{-2}$, $M_{disk} = 0.08$ M_{Sun}). Thus, whatever the true distribution of disks within Fig. 2—unless it just happens that disks with $\tau_{giant} \sim \tau_{disk}$ are somehow preferred (*29*)—it is likely that only a minority will lie within this region. Furthermore, even within this subset there are large stochastic variations, as evidenced by Fig. 2; in only one of the outcomes ($\alpha = 3 \times 10^{-3}$, $M_{disk} = 0.05$ M_{Sun}) do the gas giants bear a reasonable resemblance to Jupiter and Saturn (for an animated example, see movie S3). The other potential pathway to a solar system analog are cases in which all gas giants except two are engulfed by the star; however, because both survivors must somehow themselves undergo little migration, such outcomes also appear improbable (Fig. 2 shows no

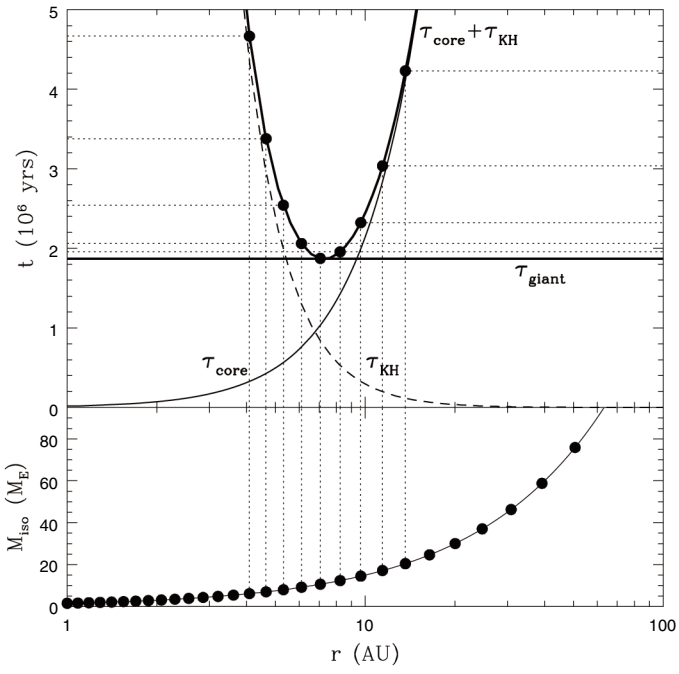


Fig. 3. Approximate timing and location of gas giant formation in a protoplanetary disk. (**Bottom**) The final or "isolation" mass of solid cores (black dots), with spacing between successive cores taken from planet formation simulations (*38*). (**Top**) The time (thick solid curve) for a core (black dots; vertical dotted lines connect to corresponding core in bottom panel) to become a gas giant (horizontal dotted lines show times for individual protoplanets). We approximate this as the sum of the time for the core to reach its final mass, τ_{core} (thin solid curve), and the time for the core to undergo runaway gas accretion, taken to be its Kelvin-Helmholtz time (*39*), τ_{KH} (dashed curve). As in more detailed calculations (*40*), we find that gas giant formation commences at one particular radius, which for typical parameters lies in or near the Jupiter-Saturn region (in this case at 7 AU and at time τ_{giant} just under 2 My). Giant formation begins in a burst, with several planets growing in rapid succession, then slows down as it spreads to larger and smaller radii. In practice, once an inner hole forms in the gas disk, formation is constrained to progress only outwards.

candidates for this scenario). All of this leads us to predict that within the diverse ensemble of planetary systems, ones resembling our own are the exception rather than the rule. Observations may be hinting at this already (*30*), although the true planet distribution remains largely obscured by selection effects (*25*). On the other hand, scaled-down versions of the solar system, in which a moderate amount of migration took place, are likely to be more common; indeed, such a system has recently been discovered through microlensing (*31*). Finally, scenarios in which type II migration is reduced (*32*, *33*) would modify our prediction, permitting a more common occurrence of solar system analogs.

In all of our simulations, the formation of a gas giant brings with it violent scattering of neighboring smaller bodies, including other cores about to undergo runaway gas accretion themselves. Such scattering has been proposed as the origin of Uranus and Neptune (*34*), with dynamical friction from the remnant outer planetesimal disk (not modeled here) serving to prevent their ejection and ultimately recircularize their orbits. Thus, whether or not Jupiter and Saturn analogs are rare, it is likely that Uranus and Neptune analogs are quite common.

References and Notes

1. K. E. Haisch, E. A. Lada, C. J. Lada, *Astrophys. J.* **553**, L153 (2001).
2. M. Juric, S. Tremaine, *ArXiv Astrophys.*, http://arxiv.org/abs/astroph/0703160 (2007).
3. S. Chatterjee, E. B. Ford, S. Matsumura, F. A. Rasio, *ArXiv Astrophys.*, http://arxiv.org/abs/astroph/0703166 (2007).
4. S. Udry, D. Fischer, D. Queloz, *Protostars and Planets V*, B. Reipurth, D. Jewitt, K. Keil, eds. (2007), pp. 685–699.
5. M. H. Lee, S. J. Peale, *Astrophys. J.* **567**, 596 (2002).
6. P. Goldreich, S. Tremaine, *Astrophys. J.* **241**, 425 (1980).
7. D. N. C. Lin, J. Papaloizou, *Astrophys. J.* **309**, 846 (1986).
8. W. R. Ward, *Icarus* **126**, 261 (1997).
9. S. Ida, D. N. C. Lin, *Astrophys. J.* **604**, 388 (2004).
10. Y. Alibert, C. Mordasini, W. Benz, C. Winisdoerffer, *Astron. Astrophys.* **434**, 343 (2005).
11. J. E. Chambers, *Astrophys. J.* **652**, L133 (2006).
12. S. Ida, D. N. C. Lin, *Astrophys. J.* **673**, 487 (2008).
13. F. C. Adams, G. Laughlin, *Icarus* **163**, 290 (2003).
14. A. T. Lee, E. W. Thommes, F. A. Rasio, *ArXiv Astrophys.*, http://arxiv.org/abs/0801.1926v1 (2008).
15. E. W. Thommes, *Astrophys. J.* **626**, 1033 (2005).
16. M. J. Duncan, H. F. Levison, M. H. Lee, *Astron. J.* **116**, 2067 (1998).
17. E. W. Thommes, L. Nilsson, N. Murray, *Astrophys. J.* **656**, L25 (2007).
18. C. D. Murray, S. F. Dermott, *Solar System Dynamics* (Cambridge Univ. Press, Cambridge, 1999).
19. D. J. Hollenbach, H. W. Yorke, D. Johnstone, *Protostars and Planets IV* (Univ. of Arizona Press, Tucson, AZ, 2000).
20. L. Hartmann, N. Calvet, E. Gullbring, P. D'Alessio, *Astrophys. J.* **495**, 385 (1998).
21. N. I. Shakura, R. A. Syunyaev, *Astron. Astrophys.* **24**, 337 (1973).
22. D. N. C. Lin, P. Bodenheimer, D. C. Richardson, *Nature* **380**, 606 (1996).
23. M. J. Kuchner, M. Lecar, *Astrophys. J.* **574**, L87 (2002).
24. C. Terquem, J. C. B. Papaloizou, *Astrophys. J.* **654**, 1110 (2007).
25. G. Marcy *et al.*, *Prog. Theor. Phys.* **158** (suppl.), 24 (2005).
26. D. A. Fischer, J. A. Valenti, *Scientific Frontiers in Research on Extrasolar Planets*, D. Deming, S. Seager, Eds. (2003), vol. 294 of *Astronomical Society of the Pacific Conference Series*, pp. 117–128.
27. N. C. Santos, G. Israelian, M. Mayor, *Astron. Astrophys.* **415**, 1153 (2004).
28. S. Udry, N. C. Santos, *Annu. Rev. Astron. Astrophys.* **45**, 397 (2007).
29. Scenarios have been proposed in which planets themselves are the source of viscosity (*35*, *36*). Although this would cause the two time scales to be correlated, τ_{giant} would systematically be less than τ_{disk}, thus making substantial migration likely in all planetary systems.
30. M. E. Beer, A. R. King, M. Livio, J. E. Pringle, *Mon. Not. R. Astron. Soc.* **354**, 763 (2004).
31. B. S. Gaudi *et al.*, *Science* **319**, 927 (2008).
32. A. Morbidelli, A. Crida, *Icarus* **191**, 158 (2007).
33. S. Ida, D. N. C. Lin, *ArXiv Astrophys.*, http://arxiv.org/abs/0802.1114 (2008).
34. E. W. Thommes, M. J. Duncan, H. F. Levison, *Nature* **402**, 635 (1999).
35. J. Goodman, R. R. Rafikov, *Astrophys. J.* **552**, 793 (2001).
36. R. Sari, P. Goldreich, *Astrophys. J.* **606**, L77 (2004).
37. C. Hayashi, *Prog. Theor. Phys.* **70**, 35 (1981).
38. E. Kokubo, S. Ida, *Icarus* **131**, 171 (1998).
39. G. Bryden, D. N. C. Lin, S. Ida, *Astrophys. J.* **544**, 481 (2000).
40. M. Ikoma, K. Nakazawa, H. Emori, *Astrophys. J.* **537**, 1013 (2000).

41. We thank the Spitzer Space Telescope Cycle-4 Theoretical Research Program, NSF, and the Natural Sciences and Engineering Research Council of Canada for supporting this work, and two anonymous referees for valuable comments.

28 April 2008; accepted 1 July 2008
10.1126/science.1159723

Suppressing Spin Qubit Dephasing by Nuclear State Preparation

D. J. Reilly,[1] J. M. Taylor,[2] J. R. Petta,[3] C. M. Marcus,[1]* M. P. Hanson,[4] A. C. Gossard[4]

Coherent spin states in semiconductor quantum dots offer promise as electrically controllable quantum bits (qubits) with scalable fabrication. For few-electron quantum dots made from gallium arsenide (GaAs), fluctuating nuclear spins in the host lattice are the dominant source of spin decoherence. We report a method of preparing the nuclear spin environment that suppresses the relevant component of nuclear spin fluctuations below its equilibrium value by a factor of ~70, extending the inhomogeneous dephasing time for the two-electron spin state beyond 1 microsecond. The nuclear state can be readily prepared by electrical gate manipulation and persists for more than 10 seconds.

Quantum information processing requires the realization of interconnected, controllable quantum two-level systems (qubits) that are sufficiently isolated from their environment that quantum coherence can be maintained for much longer than the characteristic operation time. Electron spins in quantum dots are an appealing candidate system for this application, as the spin of the electron is typically only weakly coupled to the environment relative to the

charge degree of freedom (*1*). Logical qubits formed from pairs of spins provide additional immunity from collective dephasing, forming a dynamical decoherence-free subspace (*2*, *3*).

Implementing any spin-qubit architecture requires the manipulation (*4*–*6*) and detection (*7*, *8*) of few-electron spin states, as yet demonstrated only in III-V semiconductor heterostructure devices such as gallium arsenide (GaAs), which in all cases comprise atoms with nonzero nuclear spin. The nuclear spins of the host lattice couple to electrons via the hyperfine interaction and causes rapid electron spin dephasing. In the GaAs devices presented here, for instance, an ensemble of initialized spin pairs will retain their phase relationship for $T_2^* \sim 15$ ns, consistent with theoretical estimates (*9*–*11*) and previous measurements (*4*). The time T_2^* represents an inhomogeneous dephasing time and can be extended with the use of spin-echo methods (*4*). Nonetheless, extending T_2^* by nuclear state preparation considerably reduces the burden of using complex pulse sequences or large field gradients to overcome the influence of fluctuating hyperfine fields when controlling spin qubits.

Proposals to reduce dephasing by nuclear state preparation include complete nuclear polarization (*12*), state-narrowing of the nuclear distribution (*12*–*15*), and schemes for decoupling the bath dynamics from the coherent evolution of the electron spin through the use of control pulses (*16*–*18*). These approaches remain largely unexplored experimentally, although recent optical experiments (*19*) have demonstrated a suppression of nuclear fluctuations in ensembles of self-assembled quantum dots.

We demonstrate a nuclear state preparation scheme in a double quantum-dot system, using an electron-nuclear flip-flop pumping cycle controlled by voltages applied to electrostatic gates. Cyclic evolution of the two-electron state through the resonance between the singlet (S) and m_s = 1 triplet (T_+) (*20*), in the presence of a small (few mT) applied magnetic field, leads to a factor of 70 suppression of fluctuations below thermal equilibrium of the hyperfine field gradient between the dots along the total field direction. It is this component of the hyperfine field gradient that is responsible for dephasing of the two-electron spin qubit formed by S and $m_s = 0$ triplet (T_0) states (*4*). Consequently, although the flip-flop cycle generates only a modest net nuclear polarization (<1%), the resulting nuclear state extends T_2^* of the S-T_0 qubit from 15 ns to beyond 1 μs. Once prepared, this nonequilibrium nuclear state persists for ~15 s, eventually recovering equilibrium fluctuations on the same time scale as the relaxation of the small induced nuclear polarization. This recovery time is longer than typical gate operation times by ~9 to 10 orders of magnitude. We propose that occasional nuclear state preparation by these methods may provide a remedy to hyperfine-mediated spin dephasing in networks of interconnected spin qubits.

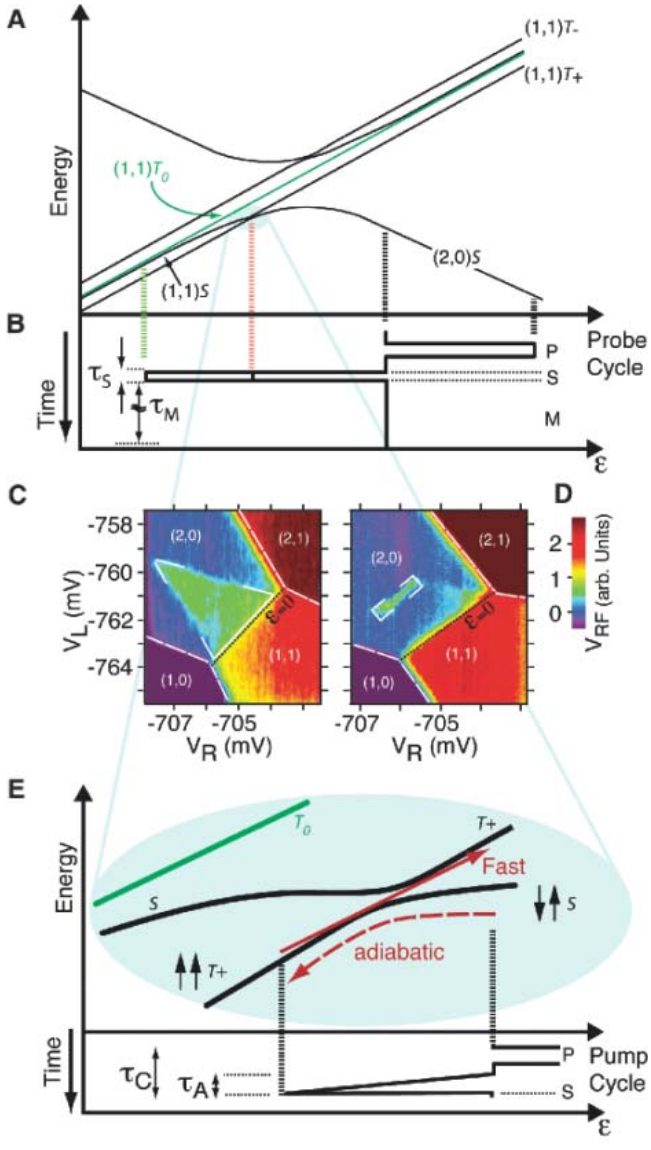


Fig. 1. (**A**) Schematic of the energy levels of the two-electron system in a magnetic field. Detuning, ε, from the (2,0)-(1,1) charge degeneracy is gate-controlled. (**B**) Gate-pulse sequence used to separately probe the longitudinal ($\Delta B_n^{\parallel}$, green dashed line) and transverse ($\Delta B_n^{\perp}$, red dashed line) components of the Overhauser field difference, depending on the position of the separation point S. (**C** and **D**) Time-averaged charge-sensing signal V_{rf} from the rf-QPC as a function of gate voltages V_L and V_R, showing features corresponding to the singlet mixing with T_0 [bracketed green triangle in (C)] and T_+ [bracketed green line segment in (D)]. (**E**) Schematic view of the S-T_+ anticrossing, illustrating the pumping cycle. With each iteration of this cycle, with period $\tau_C = 250$ ns, a new singlet state is taken adiabatically through the S-T_+ anticrossing in a time $\tau_A = 50$ ns, then returned nonadiabatically to (2,0) in ~1 ns, where the S state is then reloaded.

The double quantum dot is defined in a GaAs-AlGaAs heterostructure with a two-dimensional electron gas (2DEG) 100 nm below the wafer surface (density 2×10^{15} m^{-2}, mobility 20 m^2 V^{-1} s^{-1}). Negative voltages applied to Ti-Au gates create a tunable double-well potential that is tunnel-coupled to adjacent electron reservoirs (fig. S1). A proximal radio-frequency quantum point contact (rf-QPC) senses the charge state of the double dot, measured in terms of the rectified sensor output voltage V_{rf} (*21*). Measurements were made in a dilution refrigerator at a base electron temperature of 120 mK.

A schematic energy-level diagram (Fig. 1A), with (n,m) indicating equilibrium charge occupancies of the left and right dots, shows the three (1,1) triplet states (T_+, T_0, T_-) split by a magnetic field $\mathbf{B}_0$ applied perpendicular to the 2DEG. The detuning, ε, from the (2,0)-(1,1) degeneracy is controlled by high-bandwidth gate voltage pulses. The ground state of (2,0) is a singlet, with the (2,0) triplet out of the energy range of the experiment.

Each confined electron interacts with $N \sim 10^6$ nuclei via hyperfine coupling, giving rise to a spatially and temporally fluctuating effective magnetic (Overhauser) field (*9*–*11*, *22*). In the separated (1,1) state, precession rates for the two electron spins depends on their local effective fields, which can be decomposed into an average field and a difference field. It is useful to resolve $\mathbf{B}_n = (\mathbf{B}^l_n + \mathbf{B}^r_n)/2$, the Overhauser part of the total average field, $\mathbf{B}_{tot} = \mathbf{B}_0 + \mathbf{B}_n$, into components along ($B_n^{\parallel}$) and transverse ($B_n^{\perp}$) to $\mathbf{B}_{tot}$. The difference field, due only to Overhauser contributions, is given by $\Delta\mathbf{B}_n = (\mathbf{B}^l_n - \mathbf{B}^r_n)/2$, with components along ($\Delta B_n^{\parallel}$) and transverse ($\Delta B_n^{\perp}$) to $\mathbf{B}_{tot}$. At large negative ε, where the two electrons are well separated and exchange $J(\varepsilon)$ is negligible, $\Delta B_n^{\parallel}$ sets the precession rate between

[1]Department of Physics, Harvard University, Cambridge, MA 02138, USA. [2]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Department of Physics, Princeton University, Princeton, NJ 08544, USA. [4]Materials Department, University of California, Santa Barbara, CA 93106, USA.

*To whom correspondence should be addressed. E-mail: marcus@harvard.edu

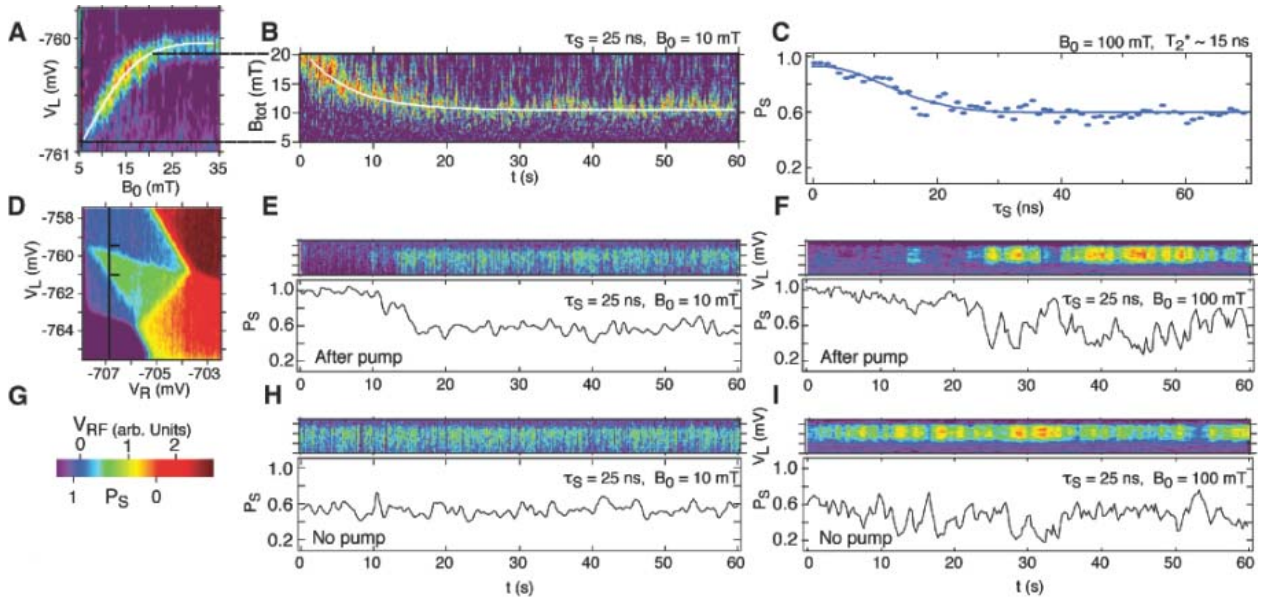


Fig. 2. (**A**) Position of the S-T_+ resonance in left-dot gate voltage V_L as a function of applied magnetic field amplitude B_0, without prior pump cycle. (**B**) Evolution of the S-T_+ position as a function of time. Vertical scale is converted from gate voltage V_L to B_{tot} via the resonance position in (A). Resonance position in gate voltage V_L and converted to B_{tot} via (A). (**C**) Singlet return probability P_S as a function of τ_S at B_0 = 100 mT. Gaussian fit gives an inhomogeneous dephasing time T_2^* = 15 ns. (**D**) Sensor output V_{rf}, as in Fig. 1C, showing triangle that yields P_S. Vertical cut (black line) with tick marks shows location of slices in upper panels of (E) to (I). (**E** and **F**) After the pump cycle, repeated slices across this triangle at the position indicated by the black line in (D) allow a calibrated measure of P_S as it evolves in time, at (E) B_0 = 10 mT and (F) B_0 = 100 mT. (**G**) Calibration between the sensing signal V_{rf} and singlet return probability P_S. (**H** and **I**) Control experiments showing slices across the triangle as in (E) and (F), but without a prior pump cycle.

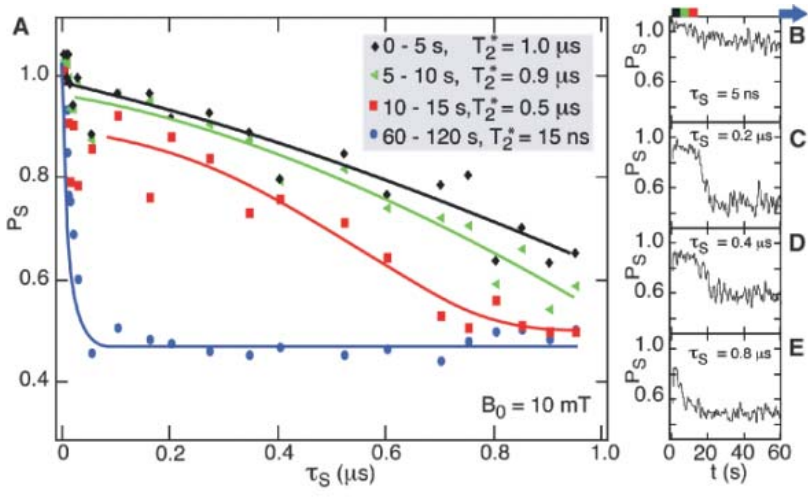


Fig. 3. (**A**) Singlet return probability P_S as a function of separation time τ_S deep in (1,1) (green dashed line in Fig. 1A), where S mixes with T_0. P_S values are shown averaged within the 0- to 5-s interval (black), the 5- to 10-s interval (green), the 10- to 15-s interval (red), and 60 to 120 s after the pump cycle (blue), along with Gaussian fits. (**B** to **E**) P_S as a function of time following the pump cycle (B_0 = 10 mT) for fixed τ_S = 5 ns (B), τ_S = 0.2 μs (C), τ_S = 0.4 μs (D), and τ_S = 0.8 μs (E).

S and T_0 states. At the value of detuning where $J(\varepsilon)$ equals the Zeeman energy $E_Z = g\mu_B B_{tot}$ (where g is the electron g factor and μ_B is the Bohr magneton), precession between S and T_+ states occurs at a rate set by $\Delta B_n^{\perp}$.

For measurement of the precession or dephasing of spin pairs in the two dots, a gate-pulse cycle ("probe cycle") first prepares (P) a singlet state in (2,0), then separates (S) the two electrons into (1,1) for a duration τ_S, then measures (M) the probability of return to (2,0). States that evolve into triplets during τ_S remain trapped in (1,1) by the Pauli blockade and are detected as such by the rf-QPC charge sensor (*4*). Figure 1, C and D, shows the time-averaged charge-sensing signal V_{rf} as a function of constant offsets to gate biases V_L and V_R, with this pulse sequence running continuously. Setting the amplitude of the S-pulse to mix S with T_0 at large detuning (green dashed line, Fig. 1B) yields the "readout triangle" indicated in Fig. 1C. Within the triangle, V_{rf} is between (2,0) and (1,1) sensing values, indicating that for some probe cycles the system becomes Pauli-blockaded in (1,1) after evolving to a triplet state. Outside this triangle, alternative spin-independent relaxation pathways circumvent the blockade (*23*). For a smaller-amplitude S-pulse (red dashed line, Fig. 1B), S mixes with T_+, also leading to partial Pauli blockade and giving the narrow resonance feature seen in Fig. 1D. The dependence of the S-T_+ resonance position on applied field B_0 serves as a calibration, mapping the gate voltage V_L (at fixed V_R) into the total effective field B_{tot}, including possible Overhauser fields (Fig. 2A). The charge-sensing signal V_{rf} is also calibrated using equilibrium (1,1) and (2,0) sensing values to give the probability $1 - P_S$ that an initialized singlet will evolve into a triplet during the separation time τ_S (Fig. 2G). A fit to $P_S(\tau_S)$ (Fig. 2C) yields (*11*, *22*, *24*) a dephasing time $T_2^* = (h/2\pi)/g\mu_B\langle\Delta B_n^{\parallel}\rangle_{rms} \sim 15$ ns, where h is Planck's constant and the subscript rms denotes a root-mean-square time-ensemble average.

We now investigate effects of the electron-nuclear flip-flop cycle ("pump cycle") (Fig. 1E). Each iteration of the pump cycle moves a singlet, prepared in (2,0), adiabatically through the S-T_+ resonance, then returns nonadiabatically to (2,0), where the state is re-initialized to a singlet by exchanging an electron with the adjacent reservoir (*20*). In principle, with each iteration of this cycle, a change in the angular momentum of the electron state occurs, with a corresponding change to the nuclear system. Iterating the pump cycle at

4 MHz creates a modest nuclear polarization on the order of 1%, as seen previously (*20*). The pump cycle was always iterated for more than 1 s, and no dependence on pumping time beyond 1 s was observed. What limits the efficiency of the pumping cycle, keeping the polarization in the few-mT regime, is not understood.

Immediately after the pump cycle, the gate voltage pattern is switched to execute one of two types of probe cycles. The first type of probe cycle starts in (2,0) and makes a short excursion into (1,1) to locate the S-T_+ resonance, allowing B_{tot} to be measured via Fig. 2A. Figure 2B shows that the nuclear polarization established by the pumping cycle relaxes over ~15 s. The second type of probe cycle starts in (2,0) and makes a long excursion deep into (1,1) to measure $P_S(\tau_S)$ where exchange is small and the S and T_0 states are mixed by $\Delta B_n^{\parallel}$. We examined $P_S(\tau_S)$ at fixed τ_S as a function of time after the end of the pump cycle by sampling V_{rf} while rastering V_L across the readout triangle. The black line in Fig. 2D shows the value of V_R, with the tick marks indicating the upper and lower limits of the rastering. Slicing through the readout triangle allows P_S to be calibrated within each slice. Remarkably, we find that $P_S(\tau_S = 25\ ns)$ remains close to unity—that is, the prepared singlet remains in the singlet state after 25 ns of separation—for ~15 s following the pump cycle (Fig. 2, E and F). Note that $\tau_S = 25$ ns exceeds by a factor of ~2 the value of T_2^* measured when not preceded by the pump cycle (Fig. 2C). The time after which P_S resumes its equilibrium behavior, with characteristic fluctuations (*24*) around an average value $P_S(\tau_S = 25\ ns) = 0.5$, is found to correspond to the time for the small (~1%) nuclear polarization to relax (Fig. 2B). Measurements of $P_S(\tau_S = 25\ ns)$ using the same probe cycle without the preceding pump cycle (Fig. 2, H and I) do not show suppressed mixing of the separated singlet state.

Measurement of $P_S(\tau_S)$ as a function of τ_S shows that T_2^* for the separated singlet can be extended from 15 ns to 1 μs, and that this enhancement lasts for several seconds following the pump cycle. These results are summarized in Fig. 3. Over a range of values of τ_S, slices through the readout triangle (as in Fig. 2E) are sampled as a function of time after pumping, calibrated using the out-of-triangle background, and averaged, giving traces such as those in Fig. 3, B to E. Gaussian fits yield T_2^* ~ 1 μs for 0 to 5 s after the pump cycle and T_2^* ~ 0.5 μs for 10 to 15 s after the pump cycle. After 60 s, no remnant effect of the pump cycle can be seen, with T_2^* returning to ~15 ns, as before the pump cycle.

The root-mean-square amplitude of longitudinal Overhauser field difference, $\langle \Delta B_n^{\parallel} \rangle_{rms} = (h/2\pi)/g\mu_B T_2^*$, is evaluated using T_2^* values within several time blocks following the pump cycle (using data from Fig. 3A). The observed increase in T_2^* following the pump cycle is thus recast in terms of a suppression of fluctuations of $\Delta B_n^{\parallel}$ (Fig. 4A). Similarly, the S-T_+ mixing rate is used to infer the size of fluctuations of the transverse component of the Overhauser field, $\langle \Delta B_n^{\perp} \rangle_{rms}$. Figure 4B shows $P_S(\tau_S = 25\ ns)$ near the S-T_+ resonance. Unlike S-T_0 mixing, which is strongly suppressed by the pump cycle, the S-T_+ resonance appears as strong as before the pump cycle. This suggests that the energy gap $E_n^{\perp}$ (Fig. 4E) is not closed by the pump cycle. Note that fluctuations in $\Delta B_n^{\perp}$ produce fluctuations in $E_n^{\perp}$, which give the S-T_+ anticrossing a width in detuning ε (Fig. 4E). Converting to a width in magnetic field via Fig. 2A gives the fluctuation amplitude $\langle \Delta B_n^{\perp} \rangle_{rms}$ following the pump cycle. Figure 4C shows a representative slice taken from Fig. 4B at the position indicated by the white dashed line. Gaussian fits to each 1-s slice yield mean positions m and widths w in the magnetic field, which fluctuate in time (Fig. 4D). The increase in w for short times ($t < 10$ s) reflects gate voltage noise amplified by the saturating conversion from gate voltage to effective field at large B_{tot} (*25*). Beyond these first few seconds, w is dominated by fluctuations of $\Delta B_n^{\perp}$ but is also sensitive to fluctuations in m that result from fluctuations of $B_n^{\parallel}$ (Fig. 4E). [For $t >$ 10 s, gate voltage noise makes a relatively small (<10%) contribution to the fluctuations.] Estimating and removing the contribution due to $B_n^{\parallel}$ (*25*) gives an estimate of $\langle \Delta B_n^{\perp} \rangle_{rms}$ as a function of time following the pump cycle. These results are summarized by comparing Fig. 4A and Fig. 4F: In contrast to the strong suppression of fluctuations in $\Delta B_n^{\parallel}$ following the pump cycle, no corresponding suppression of $\langle \Delta B_n^{\perp} \rangle_{rms}$ is observed.

Reducing the cycle rate by a factor of ~10 reduces but does not eliminate the suppression of fluctuations of $\Delta B_n^{\parallel}$ [see (*25*) for a discussion of the dependence of polarization on pump cycle rate]. Also, when the pump cycle is substituted by a cycle that rapidly brings the singlet into resonance with T_0, deep in (1,1), effectively performing multiple fast measurements of $\Delta B_n^{\parallel}$, no subsequent effect on S-T_0 mixing is observed. This demonstrates that transitions involving S and T_+, rather than S and T_0, lead to the suppression of nuclear field gradient fluctuations.

The observation that an adiabatic electron-nuclear flip-flop cycle will suppress fluctuations of the nuclear field gradient has been investigated theoretically (*26*, *27*). These models explain some, but not all, of the phenomenology described here, and it is fair to say that a complete physical picture of the effect has not yet emerged. Other nuclear preparation schemes arising from various hyperfine mechanisms, not directly related to the specific pump cycle investigated here, have also been addressed theoretically in the recent literature (*28*, *29*).

Control spin qubits in the presence of time-varying equilibrium Overhauser gradients require complex pulse sequences (*4*) or control of sizable magnetic field gradients (*2*, *30*). Suppressing fluctuations of $\Delta B_n^{\parallel}$ by a factor of ~100, as demonstrated here by means of nuclear state preparation, leads to an improvement in control fidelity on the order of 10^4, assuming typical control errors, which scale quadratically with the size of the fluctuating field at low-frequencies. We further anticipate generalizations of the present results using more than two confined spins, which would allow arbitrary gradients in nuclear fields to be created by active control of Overhauser coupling.

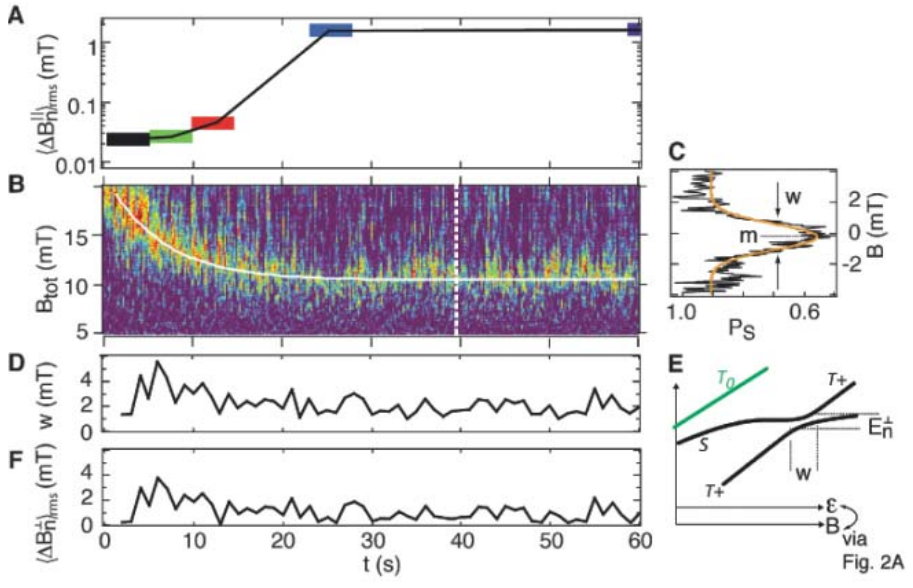


Fig. 4. (**A**) Amplitude of fluctuating longitudinal Overhauser field, $\langle \Delta B_n^{\parallel} \rangle_{rms}$, extracted from T_2^* values at the 5-s intervals in Fig. 3A. (**B**) S-T_+ resonance probed immediately after the pump cycle. Position of the resonance yields B_{tot} and its intensity gives P_S (B_0 = 10 mT, τ_S = 25 ns). (**C**) Slice from (B) at position marked by white dashed line, averaged for 1 s. For each slice a Gaussian fit yields the mean position m and width w of the S-T_+ resonance, given in units of magnetic field via Fig. 2A. (**D**) Resonance width w as a function of time after the pumping cycle. (**E**) Schematic of the S-T_+ anticrossing, showing how fluctuations of $E_n^{\perp}$ due to $\Delta B_n^{\perp}$ give the resonance a width. (**F**) Fluctuations in $\Delta B_n^{\perp}$ in terms of $\langle \Delta B_n^{\perp} \rangle_{rms}$ in a 1-s slice.

References and Notes

1. D. Loss, D. DiVincenzo, *Phys. Rev. A* **57**, 120 (1998).
2. J. Levy, *Phys. Rev. Lett.* **89**, 147902 (2002).
3. J. M. Taylor *et al.*, *Nat. Phys.* **1**, 177 (2005).
4. J. R. Petta *et al.*, *Science* **309**, 2180 (2005); published online 1 September 2005 (10.1126/science.1116955).
5. F. H. L. Koppens *et al.*, *Nature* **442**, 766 (2006).
6. K. C. Nowack, F. H. L. Koppens, Yu. V. Nazarov, L. M. K. Vandersypen, *Science* **318**, 1430 (2007); published online 1 November 2007 (10.1126/science.1148092).
7. J. M. Elzerman *et al.*, *Nature* **430**, 431 (2004).
8. S. Amasha *et al.*, *Phys. Rev. Lett.* **100**, 046803 (2008).
9. S. I. Erlingsson, Y. V. Nazarov, V. I. Fal'ko, *Phys. Rev. B* **64**, 195306 (2001).
10. A. V. Khaetskii, D. Loss, L. Glazman, *Phys. Rev. Lett.* **88**, 186802 (2002).
11. I. A. Merkulov, A. L. Efros, M. Rosen, *Phys. Rev. B* **65**, 205309 (2002).
12. A. Imamoglu, E. Knill, P. Zoller, *Phys. Rev. Lett.* **91**, 017402 (2003).
13. D. Klauser, W. A. Coish, D. Loss, *Phys. Rev. B* **73**, 205302 (2006).
14. G. Giedke, J. M. Taylor, D. D'Alessandro, M. D. Lukin, A. Imamoglu, *Phys. Rev. A* **74**, 032316 (2006).
15. D. Stepanenko, G. Burkard, G. Giedke, A. Imamoglu, *Phys. Rev. Lett.* **96**, 136401 (2006).
16. K. Khodjasteh, D. A. Lidar, *Phys. Rev. Lett.* **95**, 180501 (2005).
17. W. Zhang, V. V. Dobrovitski, L. F. Santos, L. Viola, B. N. Harmon, *Phys. Rev. B* **75**, 201302(R) (2007).
18. W. Yao, R.-B. Liu, L. J. Sham, *Phys. Rev. Lett.* **98**, 077602 (2007).
19. A. Greilich *et al.*, *Science* **317**, 1896 (2007).
20. J. R. Petta *et al.*, *Phys. Rev. Lett.* **100**, 067601 (2008).
21. D. J. Reilly, C. M. Marcus, M. P. Hanson, A. C. Gossard, *Appl. Phys. Lett.* **91**, 162101 (2007).
22. J. M. Taylor *et al.*, *Phys. Rev. B* **76**, 035315 (2007).
23. A. C. Johnson *et al.*, *Nature* **435**, 925 (2005).
24. D. J. Reilly *et al.*, http://arxiv.org/abs/0712.4033 (2007).
25. See supporting material on *Science* Online.
26. G. Ramon, X. Hu, *Phys. Rev. B* **75**, 161301(R) (2007).
27. W. M. Witzel, S. Das Sarma, *Phys. Rev. B* **77**, 165319 (2008).
28. H. Christ, J. I. Cirac, G. Giedke, *Phys. Rev. B* **75**, 155324 (2007).
29. J. Danon, Y. V. Nazarov, *Phys. Rev. Lett.* **100**, 056603 (2008).
30. L.-A. Wu, D. A. Lidar, *Phys. Rev. Lett.* **88**, 207902 (2002).

31. We thank L. DiCarlo and A. C. Johnson for technical contributions, and W. Coish, F. Koppens, D. Loss, and A. Yacoby for useful discussions. Supported by ARO/IARPA, DARPA, and NSF-NIRT (grant EIA-0210736). Research at UCSB was supported in part by QuEST, an NSF Center.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1159221/DC1
Materials and Methods
Figs. S1 to S3
References

16 April 2008; accepted 2 July 2008
Published online 10 July 2008;
10.1126/science.1159221
Include this information when citing this paper.

Large Electrocaloric Effect in Ferroelectric Polymers Near Room Temperature

Bret Neese,[1,3] Baojin Chu,[1,3] Sheng-Guo Lu,[1] Yong Wang,[2] E. Furman,[1] Q. M. Zhang[1,2]*

Applying an electrical field to a polar polymer may induce a large change in the dipolar ordering, and if the associated entropy changes are large, they can be explored in cooling applications. With the use of the Maxwell relation between the pyroelectric coefficient and the electrocaloric effect (ECE), it was determined that a large ECE can be realized in the ferroelectric poly(vinylidene fluoride-trifluoroethylene) [P(VDF-TrFE)] copolymer at temperatures above the ferroelectric-paraelectric transition (above 70°C), where an isothermal entropy change of more than 55 joules per kilogram per kelvin degree and adiabatic temperature change of more than 12°C were observed. We further showed that a similar level of ECE near room temperature can be achieved by working with the relaxor ferroelectric polymer of P(VDF-TrFE-chlorofluoroethylene).

When an electric field is applied to a dielectric material, it will induce a change in the material's polarization. The consequent changes in the entropy and temperature of the material are referred to as the electrocaloric effect (ECE) (*1*–*7*). The ECE may provide an efficient means to realize solid-state cooling devices for a broad range of applications such as on-chip cooling and temperature regulation for sensors and electronic devices, provided that materials with large ECE can be developed. Refrigeration based on the ECE approach is more environmentally friendly and hence may also provide an alternative to the existing vapor-compression approach.

A large ECE requires a large entropy change associated with the polarization change, and the dielectric material must be capable of generating large polarization changes. These requirements make it advantageous to use ferroelectric materials for the ECE and to operate the materials in a temperature region above a ferroelectric (polarization ordered)–paraelectric (polarization-disordered) (F-P) phase transition, where large electric field–induced polarization changes can be achieved (*1*, *2*). Jona and Shirane showed that the ECE occurs in both ferroelectric and paraelectric phases, and the effect is larger in the paraelectric phase just above the F-P transition (*2*). Recently "giant" ECEs were reported in ferroelectric ceramic $Pb(Zr_{0.95}Ti_{0.05})O_3$ thin films at temperatures near and above the F-P transition (*6*). These results are also consistent with experimental results regarding the magnetocaloric effect (MCE) in ferromagnetic materials, in which a giant MCE was observed at temperatures above the ferromagnetic-paramagnetic transition (*8*–*14*).

In an ideal refrigeration cycle, the working material must absorb entropy from the load to be cooled while it is in thermal contact with the load [isothermal entropy change (ΔS)]. The material is then isolated from the load while the temperature is increased because of an increased applied electric field [adiabatic temperature change (ΔT)]. The material is then placed in thermal contact with the heat sink, and the entropy that was absorbed from the cooling load is transferred to the heat sink. The working material is then isolated from the heat sink and its temperature is reduced back to the temperature of the cooling load as the field is reduced and the process is repeated. Thus, both ΔS and ΔT are the key parameters for the ECE of a dielectric material for use in refrigeration (*1*, *2*, *8*–*12*).

Although large ΔT of ~12°C has been reported in ferroelectric ceramics at about 230°C, the ΔS [~8 J per kilogram per kelvin degree (J/(kgK)] was not as high as that observed in magnetic materials that exhibit a giant MCE near room temperature, where ΔS higher than 30 J/(kgK) was observed (*6*, *8*, *11*). An underlying reason for the small ΔS is that the small ionic displacements in these ferroelectric ceramics may not generate large entropy changes. However, the ordering and disordering of dipoles in ferroelectric polymers can result in a large entropy change. For example, the heat of the F-P transition for the ferroelectric polymer poly(vinylidene fluoride-trifluoroethylene) [P(VDF-TrFE)] [65/35 mole % (mol %)] is >20.9 kJ/kg [or $\Delta S \sim 56$ J/(kgK)], which is even greater than that of the recently reported giant MCE (*15*). Furthermore, these normal ferroelectric polymers have been converted to ferroelectric relaxors that exhibit a very slim electric displacement-electric field (D-E) loop near room temperature with large electric field–induced polarization, suggesting a giant ECE near room temperature (*16*–*18*).

We investigated the ECE in normal ferroelectric P(VDF-TrFE) (55/45 mol %) at temperatures above the F-P transition and in the relaxor ferroelectric polymer P(VDF-TrFE-chlorofluoroethylene) [P(VDF-TrFE-CFE)] at near room temperature around the dielectric constant maximum. We chose the P(VDF-TrFE) (55/45 mol %) copolymer be-

[1]Materials Research Institute, The Pennsylvania State University, University Park, PA 16802, USA. [2]Electrical Engineering Department, The Pennsylvania State University, University Park, PA 16802, USA. [3]Materials Science and Engineering Department, The Pennsylvania State University, University Park, PA 16802, USA.

*To whom correspondence should be addressed. E-mail: qxz1@psu.edu

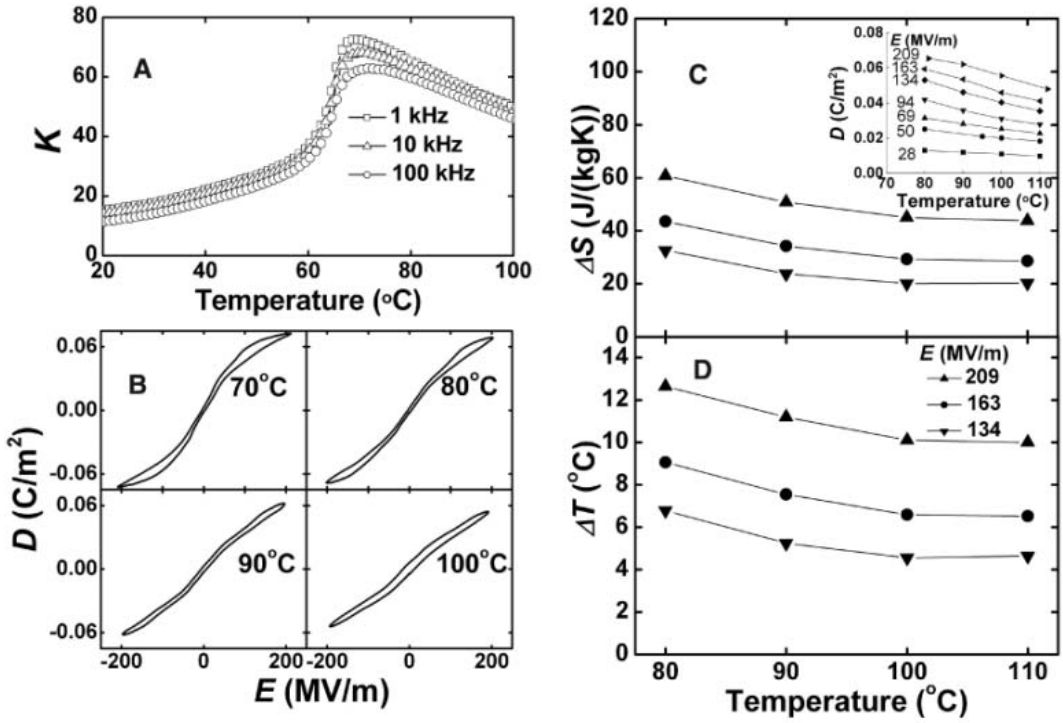


Fig. 1. (**A**) The dielectric constant of P(VDF-TrFE) 55/45 mol % copolymer films as a function of temperature measured at frequencies of 1, 10, and 100 kHz. (**B**) The D-E loops measured at 1 kHz for the copolymer films at temperatures near and above the F-P transition. (**C**) ΔS at temperatures from 80° to 110°C and different electric fields. The inset shows the displacement versus temperature taken from the D-E loops at different temperatures. (**D**) ΔT of the copolymer films.

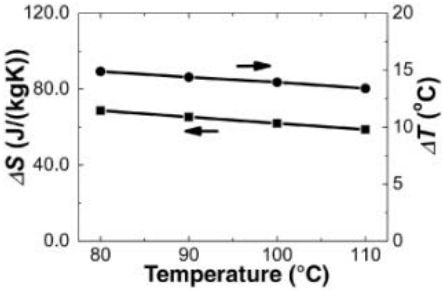


Fig. 2. ΔS and ΔT for the P(VDF-TrFE) copolymer with a 209-MV/m electric field derived from the phenomenological theory (*15*).

cause the F-P transition is continuous in it, which avoids the hysteresis effect associated with a first-order F-P transition, even at temperatures above the F-P transition (*1*, *2*, *16*, *19*). Furthermore, among all of the available P(VDF-TrFE) compositions, the polymer with 55/45 mol % exhibits the lowest F-P transition temperature (~70°C) (Fig. 1A), which is useful because we are interested in refrigerants for near-room-temperature operation (*16*, *19*).

The polymer films used in this investigation were made by the spin-cast method in the thickness range from 0.4 to 2 μm on metalized glass substrates. There was no substantial change in the ferroelectric properties in this thickness range, which is also consistent with the earlier results (*20*, *21*). The details of film preparation and characterization are presented in the supporting online material (*15*).

Figure 1B presents the D-E loops measured at 1 kHz on the P(VDF-TrFE) copolymer at temperatures at and above the F-P transition. Reversible ΔS and ΔT for a polymer film with area A as the electric field changes from E_1 to E_2 can be found from the pyroelectric coefficient as a function of the electric field (*1*–*7*)

$$\Delta S = -\int_{E_1}^{E_2} \left(\frac{\partial D}{\partial T}\right)_E dE \qquad (1a)$$

$$\text{and } \Delta T = -\frac{1}{\rho}\int_{E_1}^{E_2} \frac{T}{C}\left(\frac{\partial D}{\partial T}\right)_E dE \qquad (1b)$$

where ρ is the density and C is the specific heat. These equations are based on the Maxwell equation $(\frac{\partial D}{\partial T})_E = (\frac{\partial S}{\partial E})_T$ (*1*, *2*). For polymer films under constant mechanical stresses, the Maxwell relation should be modified to $(\frac{1}{A}\frac{\partial Q}{\partial T})_E = (\frac{\partial S}{\partial E})_T$, where Q is the charge, to take into account the secondary pyroelectric effect due to the thermally induced change in sample area (*5*, *15*, *22*). For the spin-cast films on the metalized glass substrates as investigated here, this secondary effect is negligible and $(\frac{1}{A}\frac{\partial Q}{\partial T})_E = (\frac{\partial D}{\partial T})_E$ (*15*).

The D versus T curves taken at different electric fields for the copolymer films are shown in the inset of Fig. 1C at temperatures above the F-P transition, where the copolymer displays slim polarization hysteresis loops (Fig. 1B), which reduce the effect of D-E hysteresis on the data analysis. The calculated ΔS and ΔT at several electric field levels are presented in Fig. 1, C and D, respectively. The data reveal that the copolymer exhibits a ΔS of more than 55 J/(kgK) and ΔT of more than 12°C under an electric field of 209 MV/m and temperatures around 80°C. In the entire temperature range from 80° to 110°C, ΔT is more than 10°C and ΔS is more than 45 J/(kgK). The heat capacity of the 55/45 copolymer was characterized and presented in (*15*). In deducing ΔT from Eq. 1b, the heat capacity measured at zero electric field was used. There was little change in the heat capacity at temperatures between 80° and 110°C, suggesting that it will not change greatly with the field.

The experimental data obtained are compared with the results derived on the basis of the phenomenological theory, presented in Fig. 2 (*1*, *2*, *15*, *23*, *24*). ΔS and ΔT based on the theory are higher than the experimental values by about 15% at 80°C and 25% at 110°C. Several effects can cause the difference between these. (i) The experimental films contain an amorphous phase (>25 wt %), whereas the phenomenological theory deals with the crystalline phase only. (ii) Besides the contribution from the pure crystalline phase, the conversion between the crystalline and amorphous phases will also contribute to the ECE (*25*). (iii) The ECE from the phenomenological theory is derived for the ferroelectric polymer under mechanical stress-free conditions, whereas the experimentally measured ECE is from polymer films that are subject to the mechanical constraints of the substrate.

Recent experimental results have shown that by defect modification, either through copolymerization with a bulky monomer such as CFE or CTFE (chlorotrifluoroethylene) to form a terpolymer or by direct high-energy irradiation on copolymer films, the normal ferroelectric P(VDF-TrFE) polymer can be converted into a ferroelectric relaxor, which exhibits a large room-temperature dielectric constant [~55 at 1 kHz (Fig. 3A)] (*16*–*18*). Near room temperature, these polymers exhibit a large change of D with temperature, which suggests a high ECE in these polymers near room temperature (*16*–*18*).

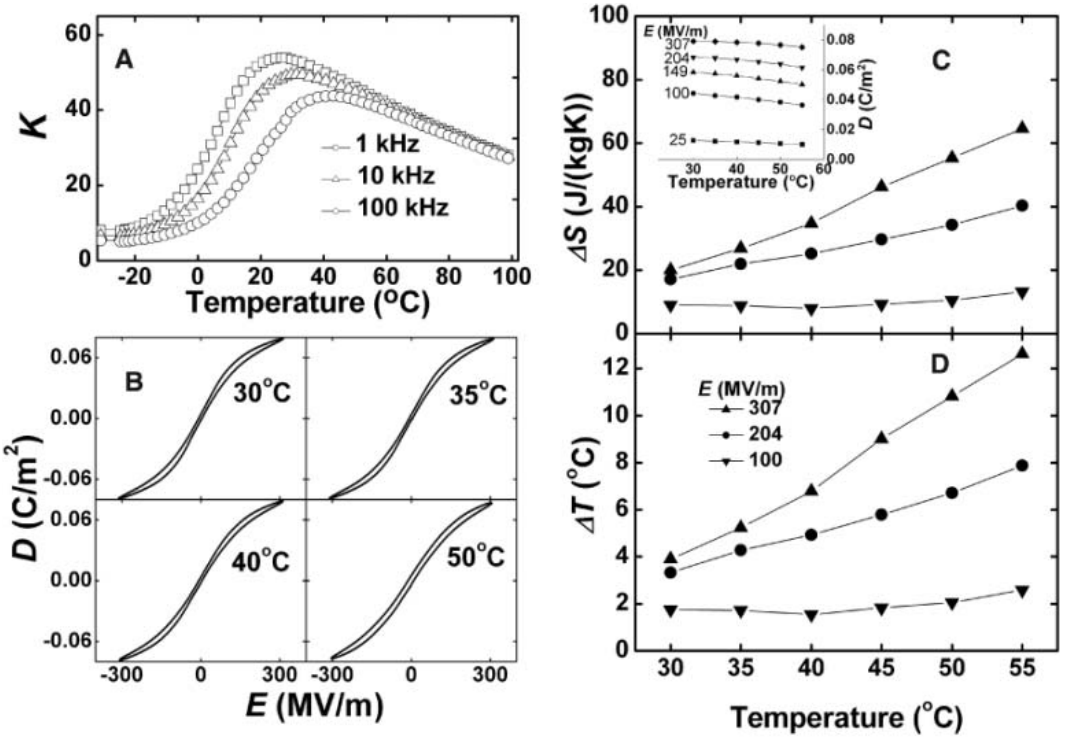


Fig. 3. (**A**) The dielectric constant of P(VDF-TrFE-CFE) 59.2/33.6/7.2 mol % terpolymer films as a function of temperature measured at frequencies of 1, 10, and 100 kHz. (**B**) The D-E loops measured at 1 kHz for the terpolymer films at temperatures near the broad dielectric constant peak. (**C**) ΔS at temperatures from 30° to 55°C and different electric fields. The inset shows the displacement versus temperature taken from the D-E loops at different temperatures. (**D**) ΔT of the terpolymer films.

The D-E loops for the P(VDF-TrFE-CFE) 59.2/33.6/7.2 mol % terpolymer are presented in Fig. 3B, from which ΔS and ΔT are determined, as shown in Fig. 3, C and D, respectively. At 55°C, $\Delta T = 12$°C and $\Delta S = 55$ J/(kgK) are derived. These values are comparable to those observed in the P(VDF-TrFE) 55/45 mol % copolymer, but a higher field is required to induce these changes. For the terpolymer, the field is ~307 MV/m, whereas for the copolymer the required field is about 209 MV/m. In the terpolymer, the defect modifications lower the peak dielectric constant, and thus in order to induce the same amount of polarization change, a higher field is required. However, because these polymers can be easily made into thin films, the applied voltage is not high (80 to 120 V for films 0.4 μm thick) to induce the observed ECE.

The data also reveal that there is a difference in the temperature dependence of ΔS and ΔT between the two ferroelectric polymer systems. The normal ferroelectric P(VDF-TrFE) copolymer displays a decrease of both ΔS and ΔT with temperature above the F-P transition, whereas in the relaxor ferroelectric P(VDF-TrFE-CFE), both ΔS and ΔT increase with temperature above the dielectric constant peak in the experimental range investigated. These results are similar to those observed in ceramic thin films of normal ferroelectric and relaxor ferroelectric material (*6*, *7*). The decrease of ΔS and ΔT with temperature is expected for a normal ferroelectric material at temperatures above the F-P transition, because a higher field is required to induce the polar phase as the temperature increases (*1*, *2*, *15*). For the relaxor ferroelectrics, the broad dielectric constant peak, whose position shifts to higher temperatures with frequencies (Fig. 3A), does not correspond to any transition (*16*–*18*, *26*). As the temperature is lowered toward the dielectric peak, the nanopolar domains will form gradually, and hence the polarization response can be from different polarization mechanisms (*16*–*18*, *26*, *27*). The pure nanodomain reorientations may not generate as much entropy change as the field-induced nonpolar-to-polar phase transformation. As temperature increases above the dielectric constant maximum, the contribution of nanopolar-domain reorientations to the polarization response will decrease, which may be responsible for the increase of ΔS and ΔT with temperature observed in the temperature range here and in (*7*).

The results presented indicate the potential of polarpolymers to achieve a high ECE because of the large entropy change associated with the electric field–induced dipole ordering-disordering (O-D) processes at temperatures near the O-D transformations. Considering the availability of a large number of polarpolymers, including liquid-crystal polymers, a large ECE from these polymers presents a very attractive approach for practical room-temperature refrigeration. Furthermore, the large ECE and MCE observed in ferroelectric materials and ferromagnetic materials also suggest an investigation of the approach of using elastic stress to induce entropy and temperature changes (elastocaloric effect), which could be significant in polymers.

References and Notes

1. M. Lines, A. Glass, *Principles and Applications of Ferroelectrics and Related Materials* (Clarendon Press, Oxford, 1977), pp. 148–150.
2. F. Jona, G. Shirane, *Ferroelectric Crystals* (Dover, New York, 1993), pp. 134–135.
3. S. Lang, *Ferroelectrics* **11**, 519 (1976).
4. B. A. Tuttle, D. A. Payne, *Ferroelectrics* **37**, 603 (1981).
5. R. A. Anderson, R. G. Kepler, *Ferroelectrics* **32**, 13 (1981).
6. A. S. Mischenko *et al.*, *Science* **311**, 1270 (2006).
7. A. S. Mischenko *et al.*, *Appl. Phys. Lett.* **89**, 242912-1-3 (2006).
8. V. Provenzano, A. J. Shapiro, R. D. Shull, *Nature* **429**, 853 (2004).
9. K. Gschneidner Jr., *et al.*, *Mater. Sci. Forum* **315-3**, 69 (1999).
10. O. Tegus *et al.*, *Nature* **415**, 150 (2002).
11. H. Wada, Y. Tanabe, *Appl. Phys. Lett.* **79**, 3302 (2001).
12. A. Giguere *et al.*, *Phys. Rev. Lett.* **83**, 2262 (1999).
13. V. K. Pecharsky *et al.*, *Phys. Rev. Lett.*, **91,** 197204-1-4 (2003).
14. D. Haskel *et al.*, *Phys. Rev. Lett.* **98**, 247025-1-4 (2007).
15. See supporting material on *Science* Online.
16. Q. M. Zhang, C. Huang, F. Xia, J. Su, in *Electroactive Polymer Actuators as Artificial Muscles*, Y. Bar-Cohen, Ed. (SPIE Optical Engineering Press, Bellingham, WA, 2004), chap 4.
17. Q. M. Zhang, V. Bharti, X. Zhao, *Science* **280**, 2101 (1998).
18. F. Xia *et al.*, *Adv. Mater.* **14**, 1574 (2002).
19. H. Nalwa, Ed., *Ferroelectric Polymers* (Marcel Dekker, New York, 1995).
20. Q. M. Zhang *et al.*, *J. Appl. Phys.* **89**, 2613 (2001).
21. F. Xia *et al.*, *Appl. Phys. Lett.* **78**, 1122 (2001).
22. T. Brossat *et al.*, *C. R. Acad. Sci. Paris* **B288**, 53 (1979).
23. S. Ducharme *et al.*, *Phys. Rev. Lett.* **84**, 175 (2000).
24. T. Furukawa, *Ferroelectrics* **57**, 63 (1984).
25. R. A. Anderson, R. G. Kepler, R. Lagasse, *Ferroelectrics* **33**, 91 (1981).
26. L. E. Cross, *Ferroelectrics* **76**, 241 (1987).
27. Z.-Y. Cheng *et al.*, *IEEE Trans. Ultrasonics Ferro. Freq. Control* **47**, 1296 (2000).

28. We thank L. E. Cross for stimulating discussions regarding this work. The work was supported by the U.S. Department of Energy through grant number DE-FG02-07ER46410.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5890/821/DC1
Materials and Methods
Figs. S1 and S2
References

25 April 2008; accepted 26 June 2008
10.1126/science.1159655

Programming DNA Tube Circumferences

Peng Yin,[1,2,3]† Rizal F. Hariadi,[4] Sudheer Sahu,[6] Harry M. T. Choi,[2] Sung Ha Park,[1,5]* Thomas H. LaBean,[6,7] John H. Reif[6]

Synthesizing molecular tubes with monodisperse, programmable circumferences is an important goal shared by nanotechnology, materials science, and supermolecular chemistry. We program molecular tube circumferences by specifying the complementarity relationships between modular domains in a 42-base single-stranded DNA motif. Single-step annealing results in the self-assembly of long tubes displaying monodisperse circumferences of 4, 5, 6, 7, 8, 10, or 20 DNA helices.

DNA, life's information carrier, has recently emerged as a versatile material for constructing self-assembled synthetic molecular structures and devices (*1*–*4*). The construction of extended DNA arrays has motivated the search for rigid molecular building blocks, as component rigidity is commonly considered necessary for the formation of well-ordered DNA crystals rather than ill-defined aggregates (*5*). A typical building block, or tile, has a rigid structural core and displays several "sticky ends" that allow for specific binding with other tiles to guide lattice formation (*6*, *7*). Diverse tiling lattices have been constructed (*8*), and some of these lattices are reported to form tubes (*9*–*15*). Such DNA nanotubes typically have varied circumferences.

As the precise control of the structure of matter is a central goal for nanotechnology, materials science, and supermolecular chemistry, controlling DNA tube circumferences has attracted intense research interest. One strategy is to encode the circumferential tube geometry directly in each individual building block (*16*–*21*): barrel-like and/or half-barrel–like rigid tiles with designed tubular curvature and circumference are first assembled and then stacked to produce tubes with prescribed circumferences. Using this strategy, researchers have successfully constructed DNA tubes containing three (*17*, *18*), six (*16*, *20*, *21*), and eight circumferential helices (*21*) and have proposed designs for tubes of arbitrary circumference (*19*). However, this approach requires the circumference-dependent construction of distinct building blocks that often have complicated molecular structures. This motivates us to search for alternative strategies that are modular and simpler.

We report the construction of DNA lattices using a flexible, single-stranded DNA motif, which is substantially simpler than the current practice of using multistranded rigid tiles. During lattice formation, the motif configures itself into a tilelike geometry, and motif-motif interactions result in emergent rigidity along the extended growth direction of the lattice. Importantly, this flexible motif allows us to program the tube circumference also as an emergent property collectively defined by the modular interactions between the motifs. In the resulting framework, simply pairing modular domains in the single motif results in the self-assembly of monodisperse DNA tubes with designed circumferences. Additionally, the motif-based, codified construction permits the description of a tube design in the form of an abstract "molecular program," further simplifying the design process.

The 42-nucleotide (nt) single-stranded DNA motif has four concatenated modular domains (Fig. 1A): The orange domain 1 and the blue domain 2 together contain 21 nucleotides; the green domain 3 and the pink domain 4 together contain 21 nucleotides. By pairing up complementary domains, the motifs can be arranged to form DNA lattices composed of parallel DNA helices connected by single-stranded linkages [or half-crossovers (*22*)] (Fig. 1B). As the orange-blue domains and the green-pink domains in a motif each measure 21 nt, the interhelix linkages are spaced periodically at every two full helical turns (i.e., 21 base pairs). In the lattices, each nonboundary motif is configured into a rectangle-like shape and is connected to four adjacent neighbors. Thus the motif implements the functionality of a tile and is termed a single-stranded tile, or SST (see fig. S1 for the comparison between a traditional rigid multistranded tile and SST). In Fig. 1B, the number k associated with a green domain indicates the number of nucleotides contained in the domain and determines the putative, approximate interhelix curvature for an unstrained lattice (e.g., not closed into tubes) through a simple formula, $k \times 34.3° - 330°$ [see supporting online material (SOM) text S1 for details].

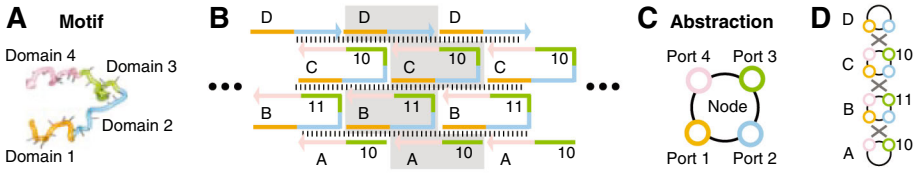


Fig. 1. (**A**) Motif. Colored lines represent modular domains; the arrowhead indicates the 3′ end. (**B**) Secondary structure of DNA lattices. Short vertical bars represent base-pairing. The shaded area indicates a repeating structural unit. (**C**) Abstraction of the motif as a node with four ports (*24*). The ports are depicted as colored circles. The color use is consistent with that in (A), and the ordering of the ports is specified by their colors: orange → blue → green → pink. (**D**) Complementarity graph.

[1]Department of Computer Science, California Institute of Technology, Pasadena, CA 91125, USA. [2]Department of Bioengineering, California Institute of Technology, Pasadena, CA 91125, USA. [3]Center for Biological Circuit Design, California Institute of Technology, Pasadena, CA 91125, USA. [4]Department of Applied Physics, California Institute of Technology, Pasadena, CA 91125, USA. [5]Center for the Physics of Information, California Institute of Technology, Pasadena, CA 91125, USA. [6]Department of Computer Science, Duke University, Durham, NC 27708, USA. [7]Department of Chemistry, Duke University, Durham, NC 27708, USA.

*Present address: Department of Physics, SKKU Advanced Institute of Nanotechnology, Sungkyunkwan University, Suwon, 440-746, Korea.

†To whom correspondence should be addressed. E-mail: py@caltech.edu

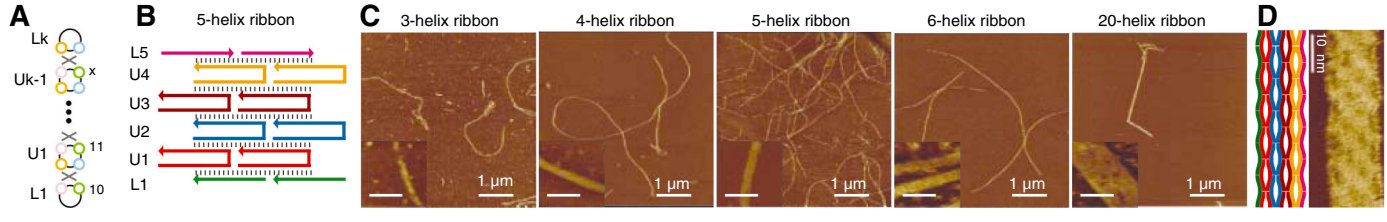


Fig. 2. Monodisperse DNA ribbons with programmed widths. (**A**) The molecular program for assembling a k-helix ribbon. (**B**) Secondary structure for the five-helix ribbon. See figs. S3 and S4 for more structures and details. (**C**) AFM images of 3-, 4-, 5-, 6-, and 20-helix ribbons. See fig. S5 for larger AFM images. (Insets) Scale bar, 50 nm. Measured ribbon widths: 9.3 ± 0.6 nm (3-helix ribbon), 12.4 ± 0.4 nm (4-helix), 15.0 ± 0.5 nm (5-helix), 18.1 ± 0.8 nm (6-helix), and 59.4 ± 1.3 nm (20-helix). See fig. S6 for ribbon width measurements. (**D**) High-resolution AFM image of the five-helix ribbon. (Left) Depiction of the expected DNA structure, emphasizing bended helices and interhelix gaps. The color use is consistent with (B). (Right) AFM image revealing an alternating pattern of four columns of interhelix gaps, in agreement with the depiction on the left. See fig. S7 for details.

The modularity and standardization of the SST motif allow us to codify the lattice design procedure. First "wire" together complementary domains and then assign the dimensions of the green domains (*23*). The codification further permits us to ignore molecular structure details and express the lattice design in a simple abstract form. Adapting a previous notation system (*24*), we abstract our motif as a node with four ports, where each port represents a modular domain (Fig. 1C). The lattice design is expressed as a complementarity graph (Fig. 1D), where two complementary ports are connected by a gray line, and the dimension of each green port is indicated with an associated number.

A complementarity graph represents a "molecular program" to be executed physically by the corresponding DNA molecules. During the execution of the program through one-pot annealing (*25*), the specified complementarity relationship between the modular domains of the motif directs the DNA molecules toward a (global or local) thermodynamic minimum on the free energy landscape, where the designed target structure resides. For example, the execution of the molecular program in Fig. 1D results in the formation of the three-helix ribbon lattice depicted in Fig. 1B.

The three-helix ribbon program can be generalized to program the formation of *k*-helix ribbons (Fig. 2A), using $(k-1)$ full SST species (U_1, U_2, ..., U_{k-1}) and two boundary half-SST species (L_1 and L_k). By executing the general program in Fig. 2A, we demonstrate the experimental construction of monodisperse ribbons (*26*) with five distinct widths: 3-, 4-, 5-, 6-, and 20-helix ribbons. The secondary structure for the five-helix ribbon is depicted in Fig. 2B. Direct imaging of the self-assembly product by atomic force microscopy (AFM) reveals the expected linear filament morphology (Fig. 2C). AFM further confirms the designed dimensions of the ribbons: a *k*-helix ribbon has a measured width of ~$3 \times k$ nm (Fig. 2C insets). Further, the morphology details of the five-helix ribbon are revealed by high-resolution AFM (Fig. 2D).

A natural strategy for constructing monodisperse *k*-helix SST tubes is to merge the two boundary half-SST species in the *k*-helix ribbon program into a full SST species (Fig. 3A). The secondary structure for $k = 6$ is described in Fig. 3B (left). The execution of the six-helix tube program through annealing results in linear filament products (Fig. 4A, third panel from left). The mechanical force exerted by repeated AFM scanning opens these filaments, which confirms their tubular nature (fig. S12). Finally, AFM width measurement of 10 random opened tubes establishes the monodispersity (i.e., no $m \times 6$ -helix tubes identified, for $m > 1$) of their circumferences (fig. S13).

This molecular implementation could, in theory, allow concatenation of multiple repeats of U1-U2-U3-U4-U5-T6 along the tube's circumference, resulting in polydisperse tubes composed of 1×6 circumferential helices, 2×6 helices, 3×6 helices, etc. Further, geometric modeling (*12*) suggests that the SST domain dimensions in Fig. 3A would result in an average interhelix curvature of ~30° per helix (SOM text S1). One would, therefore, expect 12-helix tubes to be less sterically strained (*12*) than 6-helix tubes and, thus, to dominate at thermodynamic equilibrium. The observed monodisperse formation of six-helix tubes suggests that the tube formation should be understood as a kinetic process (*10*, *15*) and that these tubes are trapped at a local minimum on the free energy landscape (Fig. 3B, right). These tubes are stable: AFM images obtained ~6 months after sample preparation revealed monodisperse six-helix tubes.

We next tested the general program (Fig. 3A) where *k* distinct SST species self-assemble into *k*-helix tubes. By choosing appropriate subsets from a common pool of 15 distinct SST species (fig. S9), we have engineered monodisperse tubes of six different circumferences: 4-, 5-, 6-, 7-, 8-, and 10-helix tubes. The generality of this strategy is further confirmed by the successful engineering of monodisperse 20-helix tubes. The secondary structures of these tubes are presented in fig. S8. Their three-dimensional (3D) illustrations are

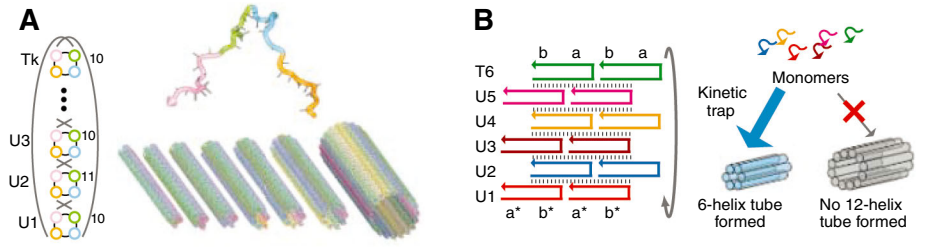


Fig. 3. Monodisperse DNA tubes with programmed circumferences. (**A**) Molecular program (left) and 3D illustrations (right) for assembling *k*-helix tubes. (**B**) Secondary structure (left) and putative kinetic trapping (right) for six-helix tubes. Asterisks denote complementarity.

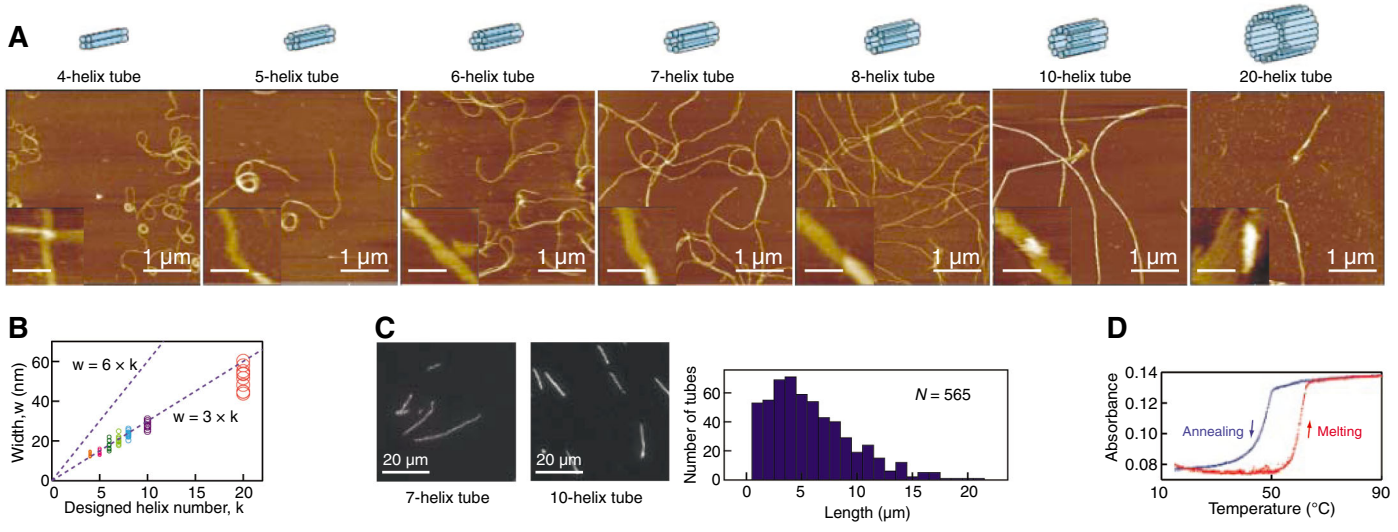


Fig. 4. (**A**) AFM images of 4-, 5-, 6-, 7-, 8-, 10-, and 20-helix tubes. See fig. S11 for larger AFM images. (Insets) Scale bar, 50 nm. (**B**) Width plot of opened tubes. A *k*-helix opened tube is expected to have a width $w \approx 3 \times k$ nm, as determined by the width measurement of the *k*-helix ribbons (fig. S6). A 2*k*-helix opened tube, in contrast, is expected to have $w \approx 6 \times k$ nm. Dashed lines corresponding to $w = 3 \times k$ and $w = 6 \times k$ are plotted to facilitate tube circumference monodispersity determination. See fig. S13 for a larger picture. (**C**) (Left) Fluorescence microscopy images of 7- and 10-helix tubes decorated with Cy3 fluorophores. (Right) Length profile of seven-helix tubes (sample size, $N = 565$). See fig. S14 for larger images and more profiles. (**D**) Annealing (blue) and melting (red) curves of four-helix tubes. Each constituent DNA strand at 100 nM. Cooling-heating rate at 0.15°C per minute. See fig. S15 for more thermal profiles.

summarized in Fig. 3A (right) and detailed in fig. S10. In each case, AFM imaging reveals the formation of long tubes (Fig. 4A), and AFM width measurement of randomly selected, opened tubes confirms the expected circumference monodispersity (Fig. 4A, insets, AFM images; Fig. 4B, a summary. See fig. S13 for details.) The length of SST tubes was investigated by using fluorescence microscopy (Fig. 4C). For seven-helix tubes, the average length is ~6 μm, with some tubes reaching ~20 μm.

Thermal formation and melting profiles of SST tubes (Fig. 4D) and SST ribbons (fig. S15) reveal hysteresis. Such hysteresis has also been observed in DNA lattices formed from multistranded tiles (*27*, *26*). It is also worth noting that the annealing and melting curves of SST tubes and ribbons demonstrate only one sharp transition temperature. This is consistent with the expectation that single-stranded DNA oligonucleotides are directly assembled into the growing lattice during annealing and disassembled from the lattice during melting. In contrast, two or more characteristic transition temperatures are commonly observed in lattices based on multistranded rigid tiles (*26*, *27*); the lowest temperature corresponds to lattice formation or melting, and the others correspond to tile formation or melting.

We suggest that the structural flexibility of SST may contribute to the success of the putative kinetic trapping of monodisperse tubes. The long sticky ends of SST and the flexible interhelix single-stranded linkage points in the assembled lattice may facilitate fast cyclization and hence trapping of the tubes with the smallest compatible number of helices. In addition, it is conceivable that in a nucleation-elongation model (*26*) (see fig. S16 for a hypothetical assembly pathway), the nucleation barrier difference between the *k*-helix tube and the 2*k*-helix tube may help trap the system into monodisperse *k*-helix tubes. The observed hysteresis (Fig. 4D) suggests the existence of a significant kinetic barrier during tube formation, and it is conceivable that this kinetic barrier is due to the presence of a nucleation barrier. It would be interesting to experimentally elucidate the kinetic assembly pathways of SST tubes. It would also be interesting to test if a similar kinetic strategy can be applied to programming the circumferences of DNA tubes assembled from multistranded rigid DNA tiles (*9*–*15*).

The ribbon and tube systems constructed here are likely to find applications ranging from biophysics to electronics and to nanotechnology. In biophysics, the programmable dimensions of the ribbons and tubes and, hence, their programmable physical properties, e.g., persistence length, make them attractive synthetic model systems. In electronics, metalization of DNA nanotubes (*9*, *11*, *17*) may result in nanowires with controlled diameters and, hence, controlled electronic properties. In nanotechnology, DNA nanotubes with programmable geometrical and mechanical properties can be used as building blocks for more sophisticated architectures and devices [e.g., tracks for molecular motors (*28*, *24*, *3*)] and as templates for organization of functional groups (*9*, *8*).

References and Notes

1. N. C. Seeman, *Nature* **421**, 427 (2003).
2. U. Feldkamp, C. M. Niemeyer, *Angew. Chem. Int. Ed.* **45**, 1856 (2006).
3. J. Bath, A. J. Turberfield, *Nat. Nanotechnol.* **2**, 275 (2007).
4. N. C. Seeman, *Mol. Biotechnol.* **37**, 246 (2007).
5. N. C. Seeman *et al.*, *Nanotechnology* **9**, 257 (1998).
6. T. J. Fu, N. C. Seeman, *Biochemistry* **32**, 3211 (1993).
7. E. Winfree, F. Liu, L. A. Wenzler, N. C. Seeman, *Nature* **394**, 539 (1998).
8. C. Lin, Y. Liu, S. Rinker, H. Yan, *ChemPhysChem* **7**, 1641 (2006).
9. H. Yan, S. H. Park, G. Finkelstein, J. H. Reif, T. H. LaBean, *Science* **301**, 1882 (2003).
10. J. C. Mitchell, J. R. Harris, J. Malo, J. Bath, A. J. Turberfield, *J. Am. Chem. Soc.* **126**, 16342 (2004).
11. D. Liu, S. H. Park, J. H. Reif, T. H. LaBean, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 717 (2004).
12. P. W. K. Rothemund *et al.*, *J. Am. Chem. Soc.* **126**, 16344 (2004).
13. D. Reishus, B. Shaw, Y. Brun, N. Chelyapov, L. Adleman, *J. Am. Chem. Soc.* **127**, 17590 (2005).
14. H. Liu, Y. Chen, Y. He, A. E. Ribbe, C. Mao, *Angew. Chem. Int. Ed.* **45**, 1942 (2006).
15. Y. Ke, Y. Liu, J. Zhang, H. Yan, *J. Am. Chem. Soc.* **128**, 4414 (2006).
16. F. Mathieu *et al.*, *Nano Lett.* **5**, 661 (2005).
17. S. H. Park *et al.*, *Nano Lett.* **5**, 693 (2005).
18. B. Wei, Y. Mi, *Biomacromolecules* **6**, 2528 (2005).
19. W. B. Sherman, N. C. Seeman, *Biophys. J.* **90**, 4546 (2006).
20. S. M. Douglas, J. J. Chou, W. M. Shih, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6644 (2007).
21. A. Kuzuya, R. Wang, R. Sha, N. C. Seeman, *Nano Lett.* **7**, 1757 (2007).
22. The half-crossover can be viewed as a simplified Holliday-junction analog, which utilizes one strand, rather than the normal two strands, at the crossover exchange point. A similar structure was previously used in constructing DNA nanotubes (*14*).
23. Due to the modularity and standardization of the motif, assigning the dimensions of all the green domains in the lattice also uniquely determines the dimensions of all the other domains.
24. P. Yin, H. M. T. Choi, C. R. Calvert, N. A. Pierce, *Nature* **451**, 318 (2008).
25. Materials and methods are available as supporting material on *Science* Online.
26. R. Schulman, E. Winfree, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15236 (2007).
27. R. D. Barish, P. W. K. Rothemund, E. Winfree, *Nano Lett.* **5**, 2586 (2005).
28. R. Pei *et al.*, *J. Am. Chem. Soc.* **128**, 12693 (2006).

29. The authors thank E. Winfree at Caltech for generously hosting the majority part of this work in his lab. For inspiring discussions, the authors thank E. Winfree, P. W. K. Rothemund, R. Schulman, V. A. Beck, J. R. Vieregg, R. D. Barish, B. Yurke, D. Y. Zhang, N. A. Pierce, S. Hamada, M. Bockrath, H. Maune, Y. Huang, C. R. Calvert, N. L. Dabby, and J. Kim. The authors are also grateful to N. A. Pierce at Caltech and J. Liu at Duke for facility support, to the Pierce group for the use of the unpublished DNA sequence design component and DNA structure illustration component of the NUPACK server (www.nupack.org), and to B. Walters for technical assistance. The fluorescence microscope was built by R.F.H. and B. Yurke. There is a patent pending on this work. This work is supported by the Center for Biological Circuit Design at Caltech, NSF grants CCF-0523555 and CCF-0432038 to J.H.R., NSF grant CBET-0508284 to T.H.L., NSF grants 0622254 and 0432193 to E. Winfree, and NSF grant 0506468 to N. A. Pierce.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5890/824/DC1
Materials and Methods
Texts S1 and S2
Figs. S1 to S16

4 March 2008; accepted 16 June 2008
10.1126/science.1157312

The Role of Excited-State Topology in Three-Body Dissociation of *sym*-Triazine

John D. Savee,[1] Vadim A. Mozhayskiy,[2] Jennifer E. Mann,[1] Anna I. Krylov,[2]* Robert E. Continetti[1]*

Molecular fragmentation into three products poses an analytical challenge to theory and experiment alike. We used translational spectroscopy and high-level ab initio calculations to explore the highly debated three-body dissociation of *sym*-triazine to three hydrogen cyanide molecules. Dissociation was induced by charge exchange between the *sym*-triazine radical cation and cesium. Calculated state energies and electronic couplings suggest that reduction initially produces a population of *sym*-triazine partitioned between the 3s Rydberg and $\pi^* \leftarrow n$ electronically excited manifolds. Analysis of the topology of these manifolds, along with momentum correlation in the dissociation products, suggests that a conical intersection of two potential energy surfaces in the 3s Rydberg manifold leads to stepwise dissociation, whereas a four-fold glancing intersection in the $\pi^* \leftarrow n$ manifold leads to a symmetric concerted reaction.

Molecular dissociation plays an important role in the chemistry of nonequilibrium environments where sufficient energy is available to break a chemical bond. Most photoinitiated dissociation processes in the lower layers of Earth's atmosphere are two-body processes, for which dynamical information can be obtained through relatively straightforward experiments. However, in combustion processes, the stratosphere, interstellar space, and other lower-density environments, high-energy processes (e.g., dissociative recombination) can read-

ily induce three-body dissociation of neutral molecules. Because of the inherent difficulties in predicting nonadiabatic behavior of excited molecules and in directly observing multiple dissociation products, the acquisition of dynamical information for three-body processes has proven to be a challenge for theoreticians and experimentalists alike.

For many of the systems for which dynamical information is available, there is a history of controversy over whether the dissociation proceeds in a stepwise or a concerted fashion (*1*). The mechanistic controversy stems from the challenges of characterizing dissociation of a polyatomic system at a molecular level. Several of the reactions are believed to proceed through symmetric transition states and involve truly simultaneous (i.e., within one rotational period) breaking of two bonds. Moreover, the reverse reaction is a termolecular association reaction and is generally improbable on account of collision statistics (*2*). Yet termolecular reactions are important; a prominent example is ozone formation, where a three-body reactive collision is an essential part of the mechanism (*3*):

$$O_2 + O + M \rightarrow O_3 + M \qquad (1)$$

$$(M \equiv O_2\ ,\ N_2)$$

sym-Triazine (Tz) is among the largest molecules known to undergo photoinduced three-body decay. Because of the relative stability of hydrogen cyanide (HCN), three-body dissociation of Tz exhibits the interesting property of producing three identical molecular products:

$$C_3N_3H_3^* \rightarrow 3HCN \qquad (2)$$

This unique feature has stirred a mechanistic controversy from past experiments where dynamical information consisted of ensemble averages rather than discrete measurement of momentum partitioning to each HCN fragment per event (*4*–*8*). In these translational spectroscopy experiments (*4*, *5*), dissociation was believed to be initiated by photoexcitation to the $\pi^* \leftarrow n$ and $\pi^* \leftarrow \pi$ electronically excited manifolds, although dissociation to 3HCN occurs ultimately through decay to the ground state, as dictated by symmetry correlation arguments and energy balance (*9*). The original study by Ondrey and Bersohn assumed that the three HCN photoproducts were produced in concerted fashion and therefore received an equal momentum partitioning (*4*). However, a later study by Gejo *et al.* presented evidence in favor of stepwise dissociation (*5*). Ab initio calculations of transition states along with classical trajectory studies have found that the barrier to a symmetric concerted reaction lies lowest in energy, but suggest that at higher excitation energies the stepwise mechanism could become more prevalent (*6*–*8*).

Here, we report translational spectroscopy of the three-body dissociation of Tz by means of coincidence detection techniques (*10*, *11*). Unlike time-of-flight techniques, which forced numerous assumptions in the previous studies, coincidence detection methods provide a complete kinematic characterization of the dissociation process. Because coincidence experiments of this nature require a fast neutral beam, Tz was excited via charge exchange (CE) between a keV Tz cation (Tz^+) beam and an alkali electron donor, cesium (Cs) (*12*, *13*). The production of excited neutral molecules by CE has been well studied (*13*–*15*), and recent studies of the CE-induced three-body dissociation of H_3 have confirmed the utility of coincidence techniques for investigating dissociation dynamics (*16*, *17*). The complex electronic structure of Tz and the intricate mechanism by which CE deposits energy in the neutral acceptor prompted a concurrent investigation using state-of-the-art ab initio methods, which proved instrumental in the interpretation of our experimental observations.

Figure 1 shows the measured kinetic energy release distribution, P(KER), for the HCN fragments produced by CE-induced dissociation of Tz with a 16-keV Tz^+ beam, accompanied by relevant dissociation limits, calculated electronic-state minima, and energy thresholds. P(KER) distributions obtained at 12 and 16 keV are similar and extend from 0 to 5 eV, with a major feature peaked at 2.6 eV, a minor feature at 0.5 eV, and a maximum KER at ~5 eV. To illustrate the KER-dependent dissociation dynamics, we divided the P(KER) distribution into 32 bins and then used Dalitz representations (*18*) to visualize the momentum partitioning among the fragments for the events contained within each bin. Several Dalitz representations constructed from the 12- and 16-keV Tz data are shown in Fig. 2, corresponding to the labeled KER ranges in Fig. 1. An "acute" feature (two slow HCN fragments and one fast HCN fragment) dominates the Dalitz representations between intervals B and E at both beam energies. A symmetric feature (equal momentum partitioning) is also present alongside the acute feature for intervals C and D at both beam energies, although it is more apparent in the 16-keV Dalitz plots. Evaluation of Dalitz plots over the full KER range reveals that the acute feature spans 1.5 to 5 eV, whereas the symmetric feature is confined in the 2- to 4-eV range (denoted in Fig. 1).

It is apparent that we have observed two unique dissociation mechanisms, one with symmetric and another with asymmetric momentum partitioning. The former is consistent with a concerted mechanism for the three-body breakup, whereas we found the latter to be consistent with stepwise decomposition (*11*). We found that the ratio of symmetric to asymmetric Dalitz features depends on the beam energy, which suggests that

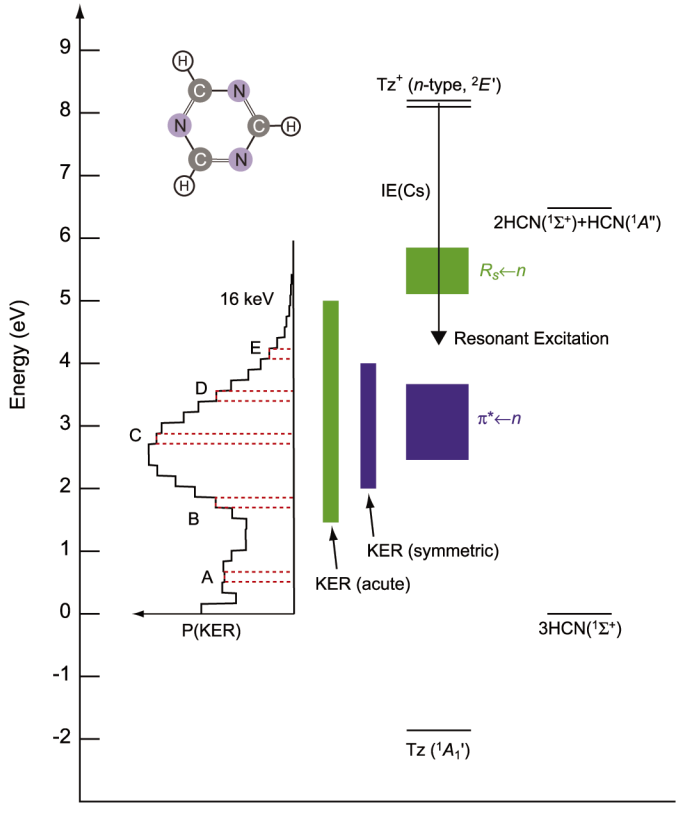


Fig. 1. Energy diagram for the three-body dissociation of Tz. The P(KER) distribution obtained with a 16-keV cation beam is shown on the left. Labeled KER intervals correspond to the following energies: A, 0.51 to 0.68 eV; B, 1.69 to 1.86 eV; C, 2.70 to 2.87 eV; D, 3.38 to 3.54 eV; E, 4.05 to 4.22 eV. The colored boxes labeled "KER (acute)" and "KER (symmetric)" mark the region over which the mechanism was observed. The colored boxes labeled "3s Rydberg" and "$\pi^* \leftarrow n$" denote the regions between the lowest- and highest-lying states (triplets included) in each manifold, as computed at the cation equilibrium geometry (C_{2v}). Zero energy corresponds to the ground-state energy of 3HCN.

[1]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093, USA. [2]Department of Chemistry, University of Southern California, Los Angeles, CA 90089, USA.

*To whom correspondence should be addressed. E-mail: krylov@usc.edu (A.I.K.); rcontinetti@ucsd.edu (R.E.C.)

different excited states are populated and consequently decay via different mechanisms.

Electronic structure calculations of Tz and $(Cs\text{-}Tz)^+$ were used to characterize the potential energy surfaces (PESs) of the relevant electronic states, as well as electronic couplings that control the CE process, allowing us to identify the gateway states for the symmetric and asymmetric dissociation and to explain the observed trends. The extremely dense electronic spectrum of Tz and extensive degeneracies (*19*–*21*) posed unique challenges for theory, which we were able to address by judicious choice of appropriate equation-of-motion coupled-cluster (EOM-CC) models (*22*–*24*) combining high accuracy with a balanced description of multiple states (*11*).

CE between keV cation beams and alkali electron donors has been shown to potentially populate multiple states in the resulting neutral species. Identifying the initially populated states that give rise to the two distinct mechanisms is a daunting task for ab initio calculations in view of the dense electronic spectrum of Tz, which consists of manifolds of valence $\pi^* \leftarrow \pi$ and $\pi^* \leftarrow n$ states (*19*–*21*), along with Rydberg states derived from the transitions from nitrogen lone pairs (n) and π orbitals. Our calculations (*11*) identified >20 states of neutral Tz below 9 eV. Energetically, CE is expected to excite Tz resonantly (i.e., at the difference between ionization energies of Tz and Cs, which are 10.01 and 3.89 eV, respectively) 6.12 eV above its ground state (*19*). However, nonresonant processes can occur when a strong perturbation acts on a zero-order system (e.g., noninteracting Cs and Tz^+) for a relatively short time, as in energetic collisions or in intense laser fields. In the context of CE, the Demkov model (*11*) demonstrates mathematically how short interacting times (i.e., large velocities) give rise to increased probability for populating nonresonant states. The total energy is conserved by redistribution of the electronic energy among internal and translational degrees of freedom of the system.

Examination of the energy diagram in Fig. 1 and computed excitation energies (*11*) reveals several singlet and triplet states that are energetically accessible according to the observed KER range (e.g., all of the $\pi^* \leftarrow n$ and $\pi^* \leftarrow \pi$ states, as well as the lowest Rydberg $R_s \leftarrow n$ state). By considering several characteristics of the excited states—their excitation energies, nonadiabatic couplings, and the topologies of the corresponding Jahn-Teller (JT) manifolds—we were able to identify the most likely initially populated states as the A_1 3s Rydberg and the A_2 $\pi^* \leftarrow n$ valence states, as follows.

To discriminate between the different states, we used a simple Demkov two-state model to evaluate transition probabilities for the CE process (*25*). The transition probability derived from this model depends on the potential energy difference (i.e., the off-resonant energy defect), the potential parameter in the coupling

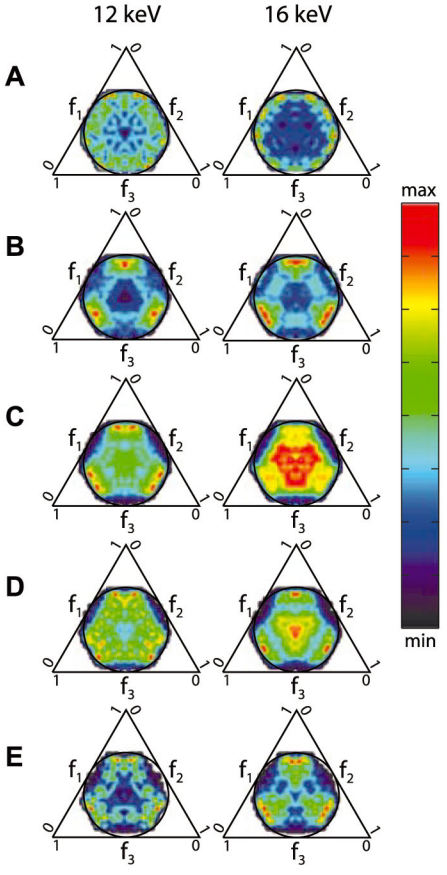


Fig. 2. Dalitz representations of the momentum correlation in the three-body breakup of Tz obtained over the KER intervals (**A** to **E**) denoted in Fig. 1. The Dalitz plot is a histogram with each of the three axes corresponding to the fractional square (f_i) of the momentum (p_i) partitioned to a particular HCN fragment ($f_i = p_i^2/\Sigma p_i^2$), and each point within the circular region of momentum conservation represents a particular arrangement of the three momentum vectors giving rise to a single event. For example, the symmetric partitioning of the momentum yields the intensity in the center of the Dalitz plots, whereas the intensity near the apexes (i.e., acute features) corresponds to one fast and two slow fragments. The Dalitz plots have been slightly cropped near the intense acute apexes to illustrate the dynamic change in other regions of the plots.

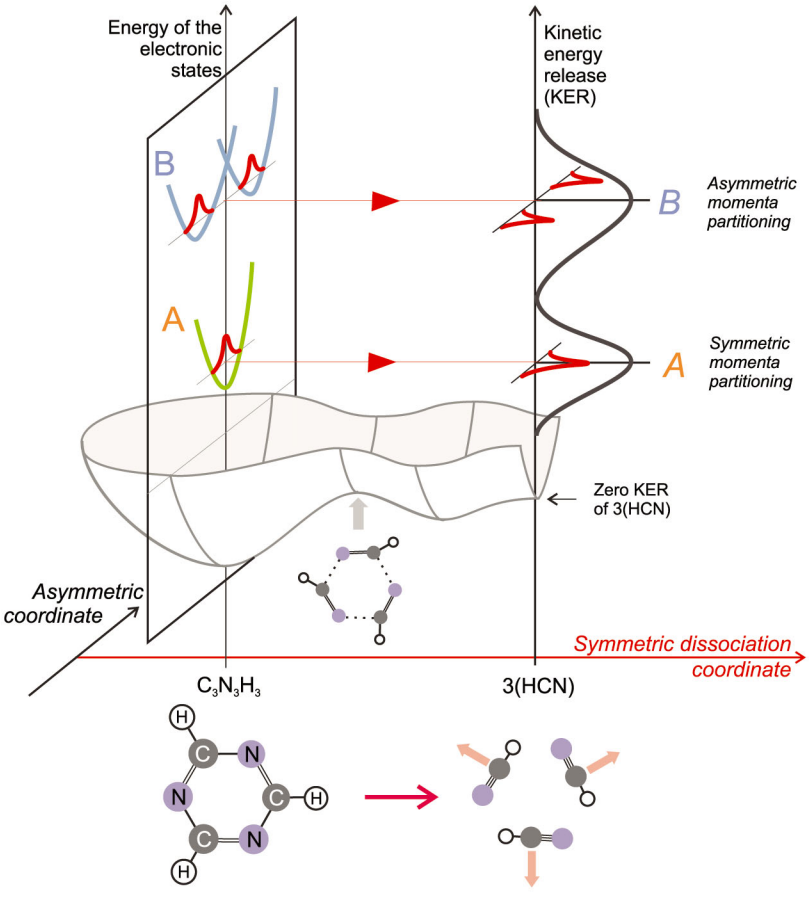


Fig. 3. Two-dimensional representation of the ground-state and excited-state PESs, demonstrating mapping of the initial wave function into the product distribution (i.e., the reflection principle). The two coordinates are the reaction coordinate for the three-body dissociation and a symmetry-lowering displacement (e.g., JT deformation). The reflection principle, which assumes ballistic dissociation on the lowest PES, predicts symmetric energy partitioning for the process initiated on the symmetric PES (right), and asymmetric partitioning for a JT distorted PES.

matrix element, and the relative velocity between the cation and the atomic electron donor (*25*). For a given interaction time determined by the atom-ion relative velocity, the model predicts a larger probability of populating states with a smaller energy defect (i.e., in resonance) and larger coupling. The velocity dependence of branching ratios between different channels is nonlinear and is determined chiefly by the energy defect. The relative probability of populating the state with a smaller energy defect (versus a larger defect) increases as the relative ion-atom velocity increases. Thus, the observed velocity-dependent change in relative intensity of the two Dalitz features qualitatively suggests that two different initial states of the neutral molecule give rise to the observed dissociation channels.

To evaluate the electronic couplings responsible for CE between Cs and Tz^+, we used EOM-CC wave functions of the combined $(Cs\text{-}Tz)^+$ system (*11*) in a generalized Mulliken-Hush (GMH) approach (*26*). The calculations show that couplings between Cs-Tz^+ and different Cs^+-Tz* states are significant for the $\pi^* \leftarrow n$ and Rydberg states. Furthermore, the $\pi^* \leftarrow \pi$ couplings are nearly zero because of the two-electron character of the corresponding electronic transition (the ground state of Tz^+ has a hole in a lone pair orbital). The same arguments apply to the corresponding triplet states, which are accessible in CE experiments. The considerable electronic coupling matrix elements for the 3s Rydberg and $\pi^* \leftarrow n$ manifolds suggest these states as the most likely initially populated electronic states. However, the coupling for the Rydberg state is two orders of magnitude larger than that of the $\pi^* \leftarrow n$ state. Thus, the Demkov model predicts a much higher probability of populating the Rydberg state, even though it is farther off-resonance than is the $\pi^* \leftarrow n$ state. The respective equilibrium geometries of these states can therefore be considered as starting points on the neutral ground-state surface by virtue of the reflection principle (Fig. 3).

A necessary condition for concerted dissociation is the existence of an appropriate energetically accessible transition state, and many theoretical studies of the interplay between concerted and stepwise mechanisms have focused on locating and comparing different transition states (*2*). Tz has an accessible symmetric transition state for three-body dissociation (*8*); however, a complete picture of the mechanism requires information about dynamics, which in turn depends on how the process was initiated. This aspect of the dynamics—the effect of initial conditions on the reaction outcome—can be understood within the reflection principle framework (*27*). In photodissociation, the reflection principle assumes a ballistic process (i.e., no vibrational equilibration) on the excited-state surface, which therefore acts as a mirror reflecting the initial wave packet onto the final product states. In the present experiment, the role of excited-state and ground-state surfaces is reversed: Tz is prepared in an electronically excited state, and dissociation occurs ultimately on the ground electronic state.

In the context of three-body breakup on the PES with a symmetric transition state, the reflection principle predicts symmetric momentum partitioning for the initially symmetric wave packet and asymmetric partitioning for the initial conditions described by the asymmetric wave packet (Fig. 3). The shape of the initial wave packet is determined by the shape of the corresponding excited-state surface on which Tz is produced by CE. Because all of the electronically excited states of Tz in the relevant energy region are derived by transitions between doubly degenerate orbitals, they are subject to asymmetric JT distortions. However, some of the states have a symmetric PES by virtue of the double degeneracy of both initial and target molecular orbitals (MOs) and can therefore result in a symmetric wave packet (*28*, *29*).

It is possible to determine which of the previously identified states have symmetric equilibrium structures and which are distorted, simply by analyzing symmetries and the electronic configurations of the corresponding wave functions, as in recent N_3^+ studies (*28*, *29*). All of the states discussed above are derived from transitions involving degenerate MOs and are therefore subject to JT distortions. States derived from the transitions between degenerate and nondegenerate MOs ($e \otimes a \rightarrow E$) form a familiar "Mexican hat"–shaped PES. The pair of $R_s \leftarrow n$ Rydberg states in Tz and the ground state of Tz^+ are of this type (*13*). A qualitatively different type of intersection occurs for the $\pi^* \leftarrow n$ states derived from excitations between two degenerate MO pairs producing two exactly degenerate and two nearly degenerate states (i.e., $e \otimes e \rightarrow E + A + B$). It can be shown that all four states are scrambled around the intersection, and linear terms for the degenerate states are very small (*28*, *29*). Consequently, the intersection appears to be glancing rather than conical. Although the minimum of the upper degenerate PES is not exactly at D_{3h}, a small-magnitude distortion (e.g., 0.001 Å and 10^{-4} eV in cyclic N_3^+) suggests a negligible effect on the corresponding nuclear wave functions, which therefore could be treated as derived from the symmetric PES.

Figure 4 shows a scan of the calculated PESs for Cs approaching Tz^+ in a direction perpendicular to the molecular plane. Dashed and solid curves correspond to the Cs-Tz^+ and Cs^+-Tz states, respectively. At infinite separation, the state ordering is exactly as in neutral Tz at the cation geometry. The pictograms on the right show the topology of each PES around D_{3h}. Note that the diffuse $R_s \leftarrow n$ states become perturbed by approaching Cs at 7 Å, whereas the PESs of the valence states remain flat up to about 3.5 Å. Analysis of the topology shows that among the states in the energy range corresponding to the symmetric dissociation, the only states with a symmetric equilibrium structure are the upper $\pi^* \leftarrow n$ states. It is therefore most likely that these states are responsible for the symmetric channel. This presents the possibility that previous experiments accessing the $\pi^* \leftarrow n$ electronic manifold were indeed observing a symmetric concerted dissociation mechanism, even though the manifold was accessed in a different manner. The asymmetric dissociation may occur via all other energetically accessible states; however, the most likely candidate is the lowest Rydberg $R_s \leftarrow n$ state for which the GMH coupling is largest. Thus, we conclude that acute and symmetric dissociation occur after initial excitation into the 3s Rydberg and $\pi^* \leftarrow n$ manifolds, respectively. This assignment is supported by qualitative agreement between the observed intensity of these channels and the GMH matrix coupling elements.

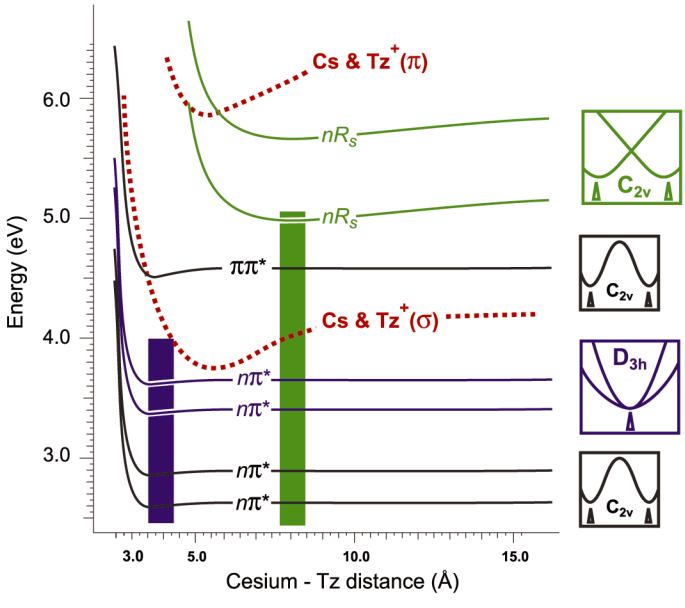


Fig. 4. Potential energy curves for the relevant singlet electronic states of the $(Cs\text{-}Tz)^+$ system in a T-shaped configuration. The geometry of the Tz moiety is that of the cation. The pictograms on the right show the PES topology for each state. As in Fig. 1, the colored boxes denote KER regions for which symmetric and asymmetric dissociation was observed. Energies are relative to the ground-state energy of 3HCN.

The energetic locations of the 3s Rydberg and $\pi^* \leftarrow n$ manifolds in the cation Franck-Condon region correlate well with the KER range over which each Dalitz feature was observed (Fig. 4). The maximum observed KER of 5 eV associated with the acute feature is close to the vertical energy for the Rydberg states (i.e., 5.17 and 5.11 eV above the 3HCN limit for the singlet and triplet, respectively). The maximum observed KER of 4 eV for symmetric dissociation lies 0.33 eV higher in energy than the vertical energy of the highest $\pi^* \leftarrow n$ state, $^1A_2$, which is located at 3.67 eV above the 3HCN limit and has D_{3h} equilibrium structure. The small separation between singlet and triplet states does not allow us to discriminate between these manifolds. However, the three-fold degeneracy of the triplets might lead one to expect triplets to be populated more frequently. On the other hand, the rate of electronic relaxation to the ground-state singlet PES is likely to be much slower for the triplets.

Nonadiabatic processes that span multiple electronic states are ubiquitous in the chemistry of energetic molecules and play a notable role in many types of natural phenomena (e.g., in biological, atmospheric, and combustion chemistry). The interactions between different electronic states are important even for reactions proceeding on a single PES, as reaction barriers can often be explained in terms of coupled diabatic states correlating with the reactants and the products. Our results show the progress being made toward analytical approaches to understanding the complexities of larger molecules.

In a broader context, this work raises a number of fundamental questions: How common is it for the equilibrium structure of an excited electronic state to determine the outcome of a reaction proceeding on another surface, and to what extent can the reflection principle be applied to polyatomic systems? With how much confidence can one rely on electronic couplings when considering the possible pathways of a nonadiabatic process? Would the difference between conical and glancing intersections manifest itself so prominently in molecules with lower symmetry? These questions all warrant future study. Our results show that synergism between theory and experiment can lead to greater understanding of how nonadiabatic behavior plays a role in these complicated systems.

References and Notes

1. C. Maul, K. H. Gericke, *Int. Rev. Phys. Chem.* **16**, 1 (1997).
2. Y. Osamura, H. F. Schaefer, M. Dupuis, W. A. Lester, *J. Chem. Phys.* **75**, 5828 (1981).
3. J. Guenther, D. Krankowsky, K. Mauersberger, *Chem. Phys. Lett.* **324**, 31 (2000).
4. G. S. Ondrey, R. Bersohn, *J. Chem. Phys.* **81**, 4517 (1984).
5. T. Gejo, J. A. Harrison, J. R. Huber, *J. Phys. Chem.* **100**, 13941 (1996).
6. Y. A. Dyakov *et al.*, *J. Phys. Chem. A* **111**, 9591 (2007).
7. J. Lee *et al.*, *Phys. Chem. Chem. Phys.* **6**, 945 (2004).
8. S. V. Pai, C. F. Chabalowski, B. M. Rice, *J. Phys. Chem.* **100**, 5681 (1996).
9. Y. Osamura, M. Unno, K. Hashimoto, *J. Am. Chem. Soc.* **109**, 1370 (1987).
10. R. E. Continetti, *Annu. Rev. Phys. Chem.* **52**, 165 (2001).
11. See supporting material on *Science* Online.
12. L. Kolaitis, D. M. Lubman, *Anal. Chem.* **58**, 1993 (1986).
13. H. Kato, K. Hirao, K. Yamashita, *J. Mol. Struct. (Theochem)* **88**, 265 (1982).
14. A. Salop, D. C. Lorents, J. R. Peterson, *J. Chem. Phys.* **54**, 1187 (1971).
15. V. Sidis, *J. Phys. Chem.* **93**, 8128 (1989).
16. C. M. Laperle, J. E. Mann, T. G. Clements, R. E. Continetti, *Phys. Rev. Lett.* **93**, 153202 (2004).
17. U. Galster, F. Baumgartner, U. Müller, H. Helm, M. Jungen, *Phys. Rev. A* **72**, 062506 (2005).
18. R. H. Dalitz, *Philos. Mag.* **44**, 1068 (1953).
19. C. Fridh, L. Asbrink, B. O. Jonsson, E. Lindholm, *Int. J. Mass Spectrom. Ion Phys.* **8**, 85 (1972).
20. K. K. Innes, I. G. Ross, W. R. Moomaw, *J. Mol. Spectrosc.* **132**, 492 (1988).
21. A. B. J. Parusel, G. Kohler, H. Nooijen, *J. Phys. Chem. A* **103**, 4056 (1999).
22. J. F. Stanton, R. J. Bartlett, *J. Chem. Phys.* **98**, 7029 (1993).
23. H. Koch, O. Christiansen, P. Jorgensen, J. Olsen, *Chem. Phys. Lett.* **244**, 75 (1995).
24. A. I. Krylov, *Annu. Rev. Phys. Chem.* **59**, 433 (2008).
25. Y. N. Demkov, *Sov. Phys. JETP* **18**, 138 (1964).
26. R. J. Cave, M. D. Newton, *Chem. Phys. Lett.* **249**, 15 (1996).
27. R. Schinke, *Photodissociation Dynamics* (Cambridge Univ. Press, Cambridge, 1993), pp. 109–133.
28. V. A. Mozhayskiy, D. Babikov, A. I. Krylov, *J. Chem. Phys.* **124**, 224309 (2006).
29. J. J. Dillon, D. R. Yarkony, *J. Chem. Phys.* **126**, 124113 (2007).

30. Supported by U.S. Air Force Office of Scientific Research grant FA9550-04-1-0035 (J.D.S., J.E.M., and R.E.C.) and NSF grant CHE-0616271 (V.M. and A.I.K.). This work was conducted under the auspices of the iOpenShell Center for Computational Studies of Electronic Structure and Spectroscopy of Open-Shell and Electronically Excited Species, supported by NSF through Chemistry Research Instrumentation and Facilities Cyberinfrastructure and Research Facilities Program (CRIF:CRF) grants CHE-0625419, CHE-0624602, and CHE-0625237.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5890/826/DC1
Materials and Methods
Figs. S1 to S10
Tables S1 to S5
References

11 March 2008; accepted 23 June 2008
10.1126/science.1157617

Phyllosilicate Diversity and Past Aqueous Activity Revealed at Mawrth Vallis, Mars

Janice L. Bishop,[1]* Eldar Z. Noe Dobrea,[2] Nancy K. McKeown,[3] Mario Parente,[4] Bethany L. Ehlmann,[5] Joseph R. Michalski,[6] Ralph E. Milliken,[2] Francois Poulet,[6] Gregg A. Swayze,[7] John F. Mustard,[5] Scott L. Murchie,[8] Jean-Pierre Bibring[6]

Observations by the Mars Reconnaissance Orbiter/Compact Reconnaissance Imaging Spectrometer for Mars in the Mawrth Vallis region show several phyllosilicate species, indicating a wide range of past aqueous activity. Iron/magnesium (Fe/Mg)–smectite is observed in light-toned outcrops that probably formed via aqueous alteration of basalt of the ancient cratered terrain. This unit is overlain by rocks rich in hydrated silica, montmorillonite, and kaolinite that may have formed via subsequent leaching of Fe and Mg through extended aqueous events or a change in aqueous chemistry. A spectral feature attributed to an Fe^{2+} phase is present in many locations in the Mawrth Vallis region at the transition from Fe/Mg-smectite to aluminum/silicon (Al/Si)–rich units. Fe^{2+}-bearing materials in terrestrial sediments are typically associated with microorganisms or changes in pH or cations and could be explained here by hydrothermal activity. The stratigraphy of Fe/Mg-smectite overlain by a ferrous phase, hydrated silica, and then Al-phyllosilicates implies a complex aqueous history.

The Mawrth Vallis outflow channel (~25°N, 20°W) cuts through the ancient cratered Noachian highland terrain near the border with the lowlands. A variety of rocks are exposed that reveal the early aqueous history of the planet. Indurated light-toned units with complex spatial and stratigraphic relations are overlain by a darker, more heavily cratered material. Erosion in many locations exposes phyllosilicate-bearing outcrop within the light-toned rocks (*1*). Here we describe the specific clay minerals identified.

Hyperspectral images acquired by the Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) (*2*, *3*) exhibit a variety of signatures in the visible/near-infrared (VNIR: ~0.4 to 4 μm) that are attributed to phyllosilicates and are consistent with observations of this area made by the Observatoire pour la Minéralogie, l'Eau, les Glaces et l'Activité (OMEGA) instrument on Mars Express (*3*–*5*). OMEGA data revealed the presence of montmorillonite and nontronite based on Al-OH and Fe-OH absorption bands near 2.2 and 2.3 μm, respectively (*4*, *6*). OMEGA data (Fig. 1) show Fe/Mg-phyllosilicates near the outflow channel and in a few other large outcrops observable at

[1]SETI Institute and NASA Ames Research Center, Mountain View, CA 94043, USA. [2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA. [3]Department of Earth and Planetary Sciences, University of California Santa Cruz, Santa Cruz, CA 95064, USA. [4]Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA. [5]Department of Geological Sciences, Brown University, Providence, RI 02912, USA. [6]Institut d'Astrophysique Spatiale, Orsay, 91405, France. [7]U.S. Geological Survey, Denver, CO 80225, USA. [8]Johns Hopkins University Applied Physics Laboratory, Laurel, MD 20723, USA.

*To whom correspondence should be addressed. E-mail: jbishop@seti.org

a resolution of hundreds of meters per pixel. Al-phyllosilicates and/or hydrated silica (not distinguished in Fig. 1) are associated with many Fe/Mg-phyllosilicate outcrops. We used the increased spatial and spectral resolution of CRISM data to more precisely identify the types of phyllosilicates present, to map them on a finer scale, and to determine their stratigraphic relationships (*7*).

CRISM spectra (Fig. 2) in the western part of Mawrth Vallis, where an areally extensive (>10 km) exposure of phyllosilicate-bearing units is observed, exhibit bands at ~1.4, 1.92, 2.30, and 2.39 μm (Fig. 2A, spectra c and d; Fig. 3B, spectra e, g, and h). The 2.30- and 2.39-μm bands lie in between the band centers observed for the Fe-smectite nontronite and the Mg-smectite hectorite (*3*, *8*) and are thus attributed to smectite having both Fe and Mg in octahedral sites such as Mg-bearing nontronite. An upward slope from ~1 to 1.5 μm is abundant in rocks surrounding the main exposure of the Fe/Mg-smectite (Fig. 2A, spectrum e; Fig. 3B, spectrum f) and is consistent with an Fe^{2+}-bearing phase such as a ferrous clay. Other spectra at higher elevations around the Fe/Mg-smectite exhibit bands near 1.4, 1.92, and 2.21 μm (Fig. 2A, spectrum g; Fig. 3A, spectrum a) that are consistent with the Al-smectite montmorillonite (*9*). Similar spectra having broad features near 2.18 to 2.26 μm

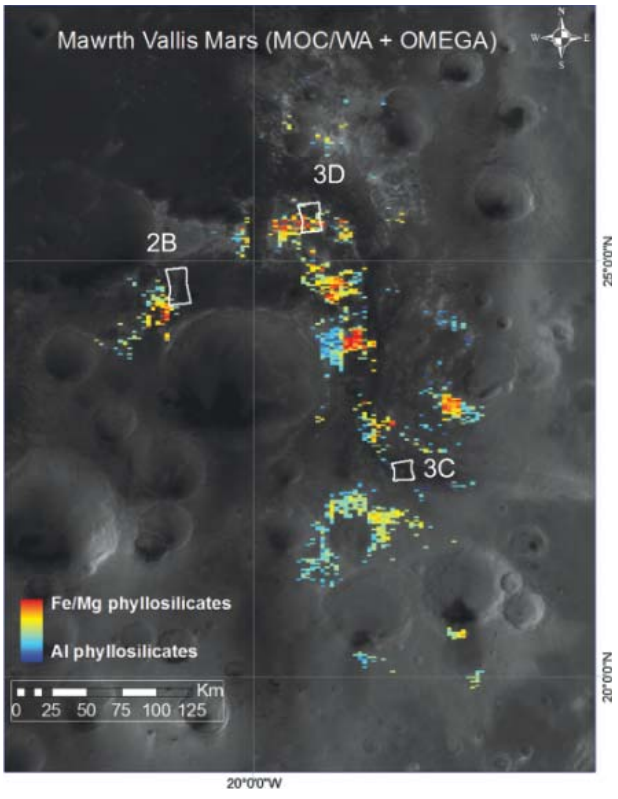


Fig. 1. Mars Orbiter Camera wide-angle (MOC/WA) image overlain with OMEGA phyllosilicate detections and locations of the CRISM images presented. Fe/Mg-smectite was mapped with the band at 2.3 μm (orange), and a combination of Al-phyllosilicate (montmorillonite/kaolinite) and hydrated silica was mapped with the 2.2-μm band (blue). The ~10-km-wide footprints are shown for three targeted CRISM images.

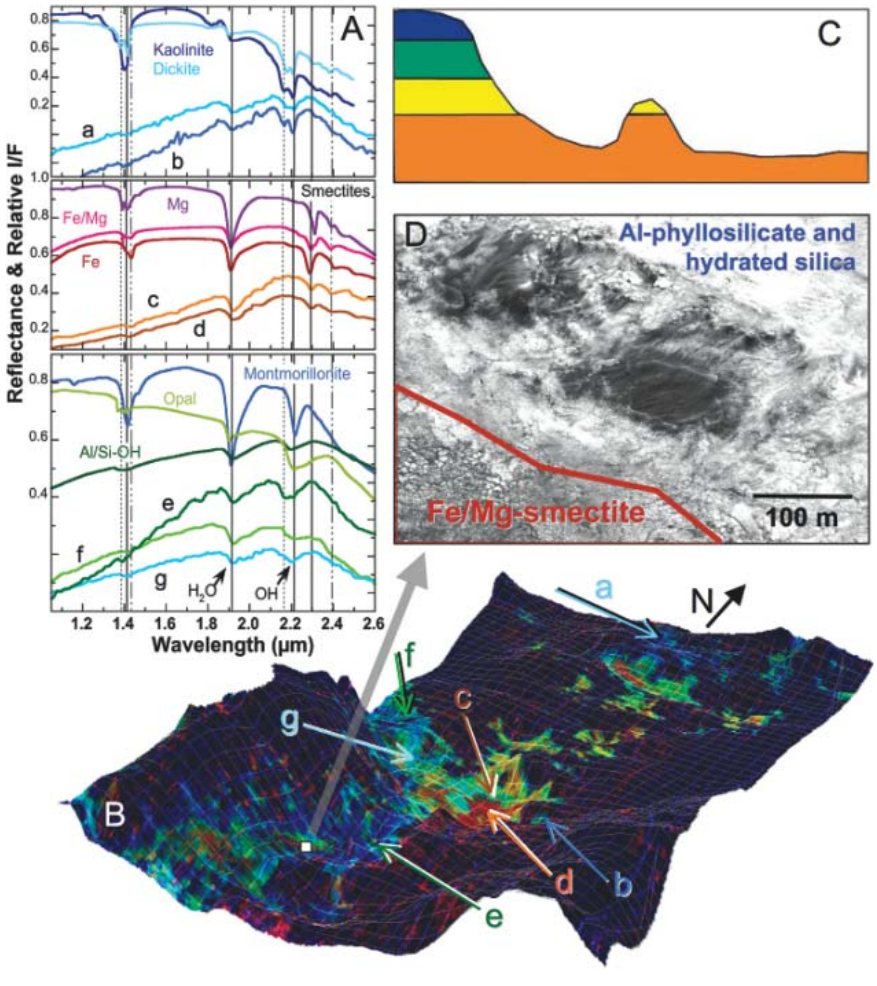


Fig. 2. Phyllosilicate-bearing deposits in western Mawrth Vallis. (**A**) CRISM spectra from image HRL000043EC (and the overlapping image FRT0000863E) compared to lab spectra of phyllosilicates and hydrated silica-bearing materials (*15*, *16*). (**B**) Mineral indicator map draped over MOLA terrain with ×20 vertical enhancement showing Fe/Mg-smectite (orange/red), Al-phyllosilicate and hydrated silica (blue), and Fe^{2+}-bearing phases (yellow/green). (**C**) Sample stratigraphy of phyllosilicates consistent with observations throughout the region, using the colors from (B). (**D**) Portion of HiRISE image PSP_005819_2050_RED, showing the transition from the polygonal Fe/Mg-smectite terrain in the lower left to the scaly Al-phyllosilicate and smooth hydrated silica in the upper right with two dark patches of mafic material.

(Fig. 2A, spectra e and f; Fig. 3A, spectra c and d) are attributed to Al/Si-OH or Si-OH (hydrated silica) (*10*). The Al-phyllosilicate and hydrated silica appear to be intermixed in most places because the sharp 1.41-, 1.91-, and 2.21-μm bands characteristic of montmorillonite are rare, whereas broad bands centered near 1.38, 1.93, and 2.21 μm (observed for opal) or 1.39, 1.92, and 2.19 μm (observed for Al/Si-OH species) are more common.

Observations from the Thermal Emission Spectrometer (with a spatial resolution of 3 to 6 km per pixel) show that the light-toned units are dominated by feldspar and a silica-rich phase (*11*). This silica detection is consistent with the Si-OH and Al/Si-OH bands identified in CRISM spectra here. Feldspar is not directly observable in the VNIR spectral region and could be present as well in the phyllosilicate-bearing rocks.

Additional spectra contain a band at ~1.4 μm and a doublet near 2.2 μm (Fig. 2A, spectra a and b), consistent with the kaolinite family. Many of these spectra also include a weak band near 1.93 μm attributed to bound water in a hydrated component (such as hydrated silica, montmorillonite, or ferrihydrite). A crisp doublet is observed at 2.16 and 2.20 μm (Fig. 2A, spectrum b) in a few cases that resembles kaolinite; in most spectra, a less resolved doublet is present at ~2.18 and 2.20 μm (Fig. 2A, spectrum a) that is more consistent with dickite.

Analyses of spectra collected at Mawrth Vallis in relation to Mars Orbiter Laser Altimeter (MOLA) topography (Figs. 2B and 3, C and D, and movie S1) reveal that Fe/Mg-smectite is prevalent in valley tributaries and other topographic lows (the southern part of the CRISM image in Fig. 3D and the central part of the image in Fig. 3C), whereas the raised knobs (Fig. 3C) and higher-elevation regions (Fig. 3D) contain Al-phyllosilicates. The Fe/Mg-smectite exposures appear to be thicker and more resistant to erosion, and in most areas they are the lowest unit visible. CRISM-MOLA analyses indicate that the Fe/Mg-smectite–rich rocks are typically 200 to 300 m lower in elevation than neighboring Al-phyllosilicate–rich and hydrated silica-rich layers and that the Al-phyllosilicates are probably covering the Fe/Mg-smectites (Fig. 2C). This is consistent with independent analyses of High Resolution Imaging Science Experiment (HiRISE) data (*12*). Many regions bordering the Fe/Mg-smectite–rich rocks (Fig. 3D) yield spectra with features similar to that of hydrated silica, plus a NIR slope that could indicate a ferrous mica such as celadonite or glauconite. However, laboratory spectra of these phyllosilicates also have weak bands near 2.2 to 2.35 μm that are not resolved in the CRISM spectra. A side view of a fissure through the channel wall (Fig. 3C) shows that Al-phyllosilicate–rich rocks lie stratigraphically above the Fe/Mg-smectite unit, which itself covers nonhydrated rocks below.

Units containing Fe/Mg-smectite, montmorillonite/hydrated silica, and kaolinite/hydrated silica exhibit distinct textures in the HiRISE images of this region (*12*). As demonstrated in Fig. 2D, the Fe/Mg-smectite unit exhibits a fractured polygonal surface consistent with the contraction of smectites exposed to dehydration. In contrast, the units containing hydrated silica possess a smooth texture, and those with montmorillonite are characterized by a polygonally fractured texture on a smaller scale than that observed for the Fe/Mg smectites. Kaolinite is observed in alteration regions as small as 50 m across, but in a few cases much larger, and appears to exhibit a smooth texture.

Phyllosilicates have been detected in a number of outcrops of Mars' ancient crust surrounding Mawrth Vallis (*13*), suggesting that the processes forming phyllosilicates here may have been widespread and that only a fraction of the phyllosilicate-rich rocks has been exposed. The initial Fe/Mg-smectite probably formed via aqueous alteration of basalt, the dominant lithology of the Martian highlands (*3*, *14*). This smectite unit appears to drape the topography and exhibits layered textures suggesting that ash was the basaltic precursor. Alteration could have occurred during direct deposition of ash onto an open body of water or by groundwater. Alternative sources include basaltic sediments and impact ejecta. The layering is complex, indicating perhaps multiple depositional events over a prolonged time period (*1*). The specific processes forming the additional clay minerals and stratigraphic relations remain unclear. Subsequent aqueous alteration and leaching of Fe and Mg from the smectite or ash could have produced hydrated silica, montmorillonite, mica, and kaolinite, as could a change in aqueous chemistry (for example, through hydrothermal activity) or alteration of a later Si-rich volcanic ash or sediment. Another possible scenario is the alteration of interbedded source rocks with different chemistries, resulting in changing alteration profiles.

Possible terrestrial analogs for the hydrated silica phases are the opal formed in hydrated ash on the southern rim of the Kilauea caldera (*15*) and hydrated amorphous Al/Si-OH phases formed in altered volcanic material at Haleakala crater in Hawaii (*16*). Acidic conditions and/or hot and humid climates support active hydrolysis and ion leaching that can result in formation of kaolinite and hydrated silica (*17*, *18*). The NIR slope interpreted as an Fe^{2+} phase is unusual because iron typically precipitates out of solution as Fe^{3+} oxides/oxyhydroxides or sulfates and because the Fe^{2+} unit occurs on top of the Fe^{3+} nontronite-like unit.

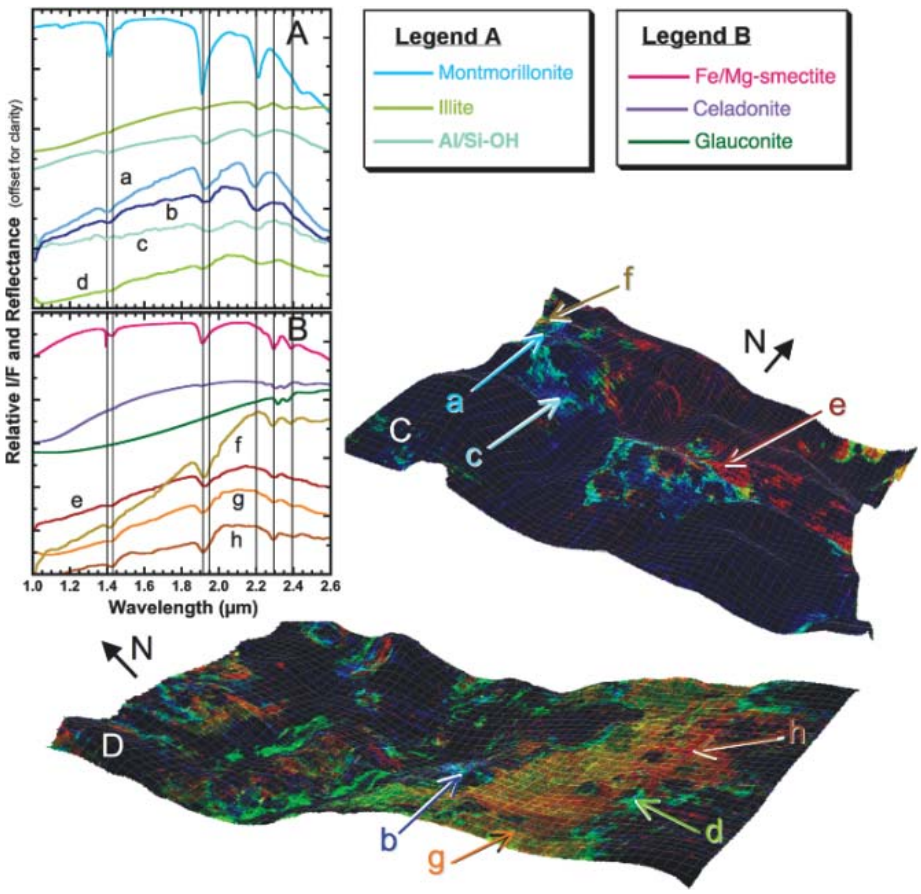


Fig. 3. Phyllosilicate-bearing deposits along the Mawrth Vallis channel. (**A** and **B**) CRISM spectra compared to lab spectra of phyllosilicates and other hydrated phases. (**C** and **D**) Mineral indicator maps of CRISM images FRT00003BFB (C) and HRL0000285A (D) draped over MOLA terrain with ×10 vertical enhancement showing Fe/Mg-smectites (orange/red), Al-phyllosilicates and hydrated silica (blue), and Fe^{2+}-bearing phases (yellow/green).

On Earth, Fe^{3+} is normally converted to Fe^{2+} in aqueous environments by acidophilic bacteria (*19*–*21*). If an Fe^{2+} phase does exist, then a reducing element must have been present at Mawrth Vallis in order to convert the Fe^{3+} to Fe^{2+}, or some form of activity [such as hydrothermal processes (*22*–*24*)] could have reduced the nontronite or released abundant Fe^{2+} in solution that precipitated out before being converted to Fe^{3+}. Thus, the complex and potentially multi-event formation conditions of phyllosilicates at Mawrth Vallis present a fascinating window into past aqueous activity on Mars.

References and Notes

1. J. R. Michalski, E. Z. Noe Dobrea, *Geology* **35**, 951 (2007).
2. S. Murchie *et al*., *J. Geophys. Res.* **112**, 10.1029/2006JE002682 (2007).
3. See supporting material on *Science* Online.
4. F. Poulet *et al*., *Nature* **438**, 623 (2005).
5. J.-P. Bibring *et al*., *Science* **307**, 1576 (2005).
6. D. Loizeau *et al*., *J. Geophys. Res.* **112**, 10.1029/2006JE002877 (2007).
7. J. F. Mustard *et al*., *Nature*, 10.1038/nature07097 (2008).
8. J. L. Bishop, E. Murad, M. D. Dyar, *Clay Miner.* **37**, 617 (2002).
9. J. L. Bishop, J. Madejova, P. Komadel, H. Fröschl, *Clay Miner.* **37**, 607 (2002).
10. R. E. Milliken *et al*., *Geology*, in press.
11. J. R. Michalski *et al*., paper presented at the 7th International Mars Conference, Pasadena, CA, 2007.
12. J. J. Wray, B. L. Ehlmann, S. Squyres, J. F. Mustard, R. L. Kirk, *Geophys. Res. Lett.*, 10.1029/2008GL034385 (2008).
13. E. Z. Noe Dobrea *et al*., paper presented at the Lunar and Planetary Science Conference, Lunar and Planetary Institute, Houston, TX, 2008.
14. J. L. Bandfield, V. E. Hamilton, P. R. Christensen, *Science* **287**, 1626 (2000).
15. J. L. Bishop, P. Schiffman, M. D. Lane, M. D. Dyar, paper presented at the Lunar and Planetary Science Conference, Lunar and Planetary Institute, Houston, TX, 2005.
16. J. L. Bishop *et al*., *Clays Clay Miner.* **55**, 1 (2007).
17. H. Chamley, *Clay Sedimentology* (Springer-Verlag, New York, 1989).
18. M. L. Jackson, *Clays Clay Miner.* **6**, 133 (1957).
19. H. A. Lowenstam, *Science* **211**, 1126 (1981).
20. K. H. Nealson, *Annu. Rev. Earth Planet. Sci.* **25**, 403 (1997).
21. T. D. Brock, J. Gustafson, *Appl. Environ. Microbiol.* **32**, 567 (1976).
22. M. J. Nelson, H. E. Newsom, D. S. Draper, *Geochim. Cosmochim. Acta* **69**, 2701 (2005).
23. J. W. Stucki, D. Tessier, *Clays Clay Miner.* **39**, 137 (1991).
24. P. R. L. Browne, *Annu. Rev. Earth Planet. Sci.* **6**, 229 (1978).

25. The authors thank E. Cloutis and an anonymous reviewer for helpful editorial comments and the CRISM Science Team for acquisition of the images. Supported by NASA's Mars Reconnaissance Orbiter project, Mars Data Analysis program, Mars Fundamental Research program, and the NASA Astrobiology Institute.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5890/830/DC1
Materials and Methods
Table S1
References
Movie S1

28 April 2008; accepted 17 June 2008
10.1126/science.1159699

Brown Carbon Spheres in East Asian Outflow and Their Optical Properties

Duncan T. L. Alexander,[1] Peter A. Crozier,[2]* James R. Anderson[3]

Atmospheric aerosols play a substantial role in climate change through radiative forcing. Combustion-produced carbonaceous particles are the main light-absorbing aerosols; thus, quantifying their optical properties is essential for determining the magnitude of direct forcing. By using the electron energy-loss spectrum in the transmission electron microscope, we quantified the optical properties of individual, submicrometer amorphous carbon spheres that are ubiquitous in East Asian–Pacific outflow. The data indicate that these common spheres are brown, not black, with a mean refractive index of $1.67 - 0.27i$ (where $i = \sqrt{1}$) at a wavelength of 550 nanometers. The results suggest that brown carbon aerosols should be explicitly included in radiative forcing models.

The importance of the impact of tropospheric aerosols on Earth's radiative balance and the uncertainty associated with radiative forcing due to aerosols are now well established by the large body of research summarized in the reports from the Intergovernmental Panel on Climate Change (*1*, *2*). A substantial fraction of the anthropogenic aerosol mass is in the form of carbonaceous particles arising from fossil fuel and biomass burning. (Carbonaceous aerosols also derive from natural biogenic emissions.) Most direct radiative forcing models classify carbonaceous particles into two main components, consisting of either negligibly absorbing organic carbon or strongly absorbing black carbon (*1*, *2*). Optical properties are incorporated through the inclusion of model parameters that account mainly for the size distribution and refractive indices of the particles.

Aside from the problem of choosing suitable refractive indices for both forms of carbon from the wide range of values found in the literature (*3*–*11*), this approach ignores the possibility of rich variations in the refractive indices of carbon arising from both changes in emission source and changes during transport in the atmosphere. To improve the accuracy of direct forcing models, it is necessary to have more precise information on the size, shape, composition, mixing state, and refractive indices of individual carbonaceous aerosols. It is difficult to determine these parameters from light-optical measurements alone because the signals are averaged over all aerosol components (size, composition, structure; see fig. S1).

We used a transmission electron microscopy (TEM) method to identify the optical properties of individual particles in ambient aerosols. We applied this method to the highly abundant carbon spheres found in all examined samples of ambient aerosols collected above the Yellow Sea during the Asian Pacific Regional Aerosol Characterization Experiment (ACE-Asia) (*12*); bulk optical properties, scanning electron microscopy images of samples taken simultaneously

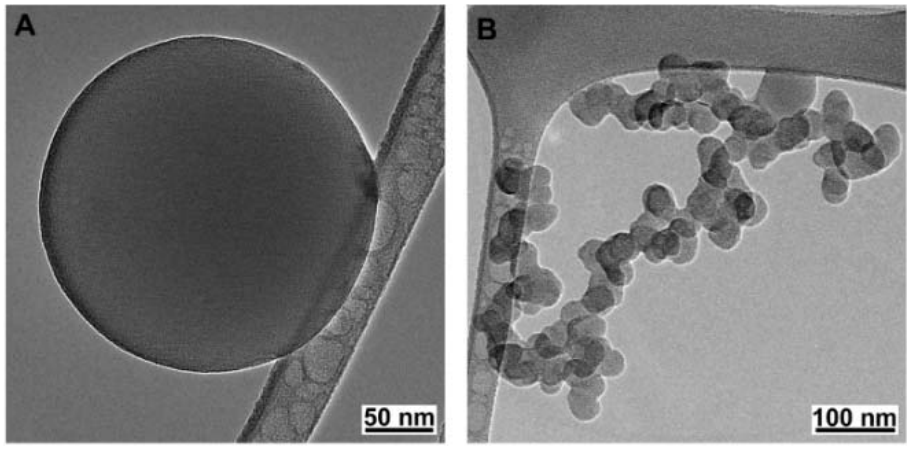


Fig. 1. A bright-field TEM image of a typical amorphous carbon sphere (**A**) and a soot carbon particle (**B**) from the atmospheric aerosol sampled at an altitude of 30 m during flight 13 of ACE-Asia.

[1]LeRoy Eyring Center for Solid State Science, Arizona State University, Post Office Box 871704, Tempe, AZ 85287, USA. [2]School of Materials, Arizona State University, Post Office Box 878706, Tempe, AZ 85287, USA. [3]Department of Mechanical and Aerospace Engineering, Arizona State University, Post Office Box 876106, Tempe, AZ 85287, USA.

*To whom correspondence should be addressed. E-mail: crozier@asi.edu

with those described here, and details of flight legs are discussed by Clarke *et al*. (*13*). The study of the spheres, and their chemical compositions, indicates that they belong to the category of "brown carbon" that has recently become the subject of intense discussion in the atmospheric community (*6*, *14*–*16*). Uniquely, we were able to specify their full-spectrum optical properties, thus providing greater insight into their impact on direct radiative forcing.

In terms of their morphology and nature, the carbon spheres differ substantially from typical soot carbon. Whereas soot carbon particles consist of aggregates of spherules mostly 20 to 60 nm in diameter, typically with an internal structure of curved graphenelike layers (*17*–*21*), and are often internally mixed with other aerosol particles, the carbon spheres are large (e.g., diameters of 100 to 400 nm), amorphous, and predominantly isolated (Fig. 1). As such, they are similar to the "tar balls" described by Pósfai *et al*. (*22*), although it is not clear that the carbon spheres in East Asian outflow described here have the same biomass-burning origin as tar balls. Further, whereas soot carbon is formed during combustion, within the flame, the carbon spheres are probably formed postcombustion by a gas-phase condensation (*22*–*24*). As a result, the spheres are always externally mixed with other carbonaceous combustion products, such as soot. The TEM method used here allows us to determine the optical properties of individual, sub-μm spheres within the externally mixed aerosol (*24*). A similar exactitude of selection is not possible with use of conventional light-optical techniques that average over the mixed aerosol (*15*). Moreover, whereas light-optical laser measurements are typically made for just one or two wavelengths, our measurements cover the full spectrum of wavelengths from the infrared, through the visible, and into the ultraviolet.

The TEM method uses high spatial resolution electron energy-loss spectroscopy (EELS) to determine the dielectric function, $\varepsilon(E) = \varepsilon_1 + i\varepsilon_2$, for a particle (*24*), which can then be transformed into refractive indices, $N = n - ik = \sqrt{\varepsilon}$, where the energy loss is $E = hc/\lambda$, with Planck's constant h, optical wavelength λ, and speed of light c (*25*). To determine optical properties in the visible regime of λ = 400 to 700 nm requires accurate EELS data from energy losses of 1.8 to 3.1 eV, which we obtained with a Tecnai F20 UT TEM (FEI, Hillsboro, Oregon) equipped with a monochromated gun, yielding a spectral resolution of 0.25 eV. The dielectric function was obtained from a Kramers-Kronig transformation of the spectrum (*26*–*28*), as described in (*24*).

Typical low-loss (0 to 10 eV) energy-loss spectra taken from the amorphous carbon spheres are shown in Fig. 2. The spectra are similar in form, indicating that the spheres share similar materials and optical properties. The π* plasmon at ~6 eV that correlates with sp^2 hybridization is not the sharp peak seen for soot carbon but rather a broad peak that extends down to ~1 eV. Figure 3A shows the energy dependence of the mean dielectric functions of 28 amorphous carbon spheres (sampled from altitudes of 30 m and 610 m), as determined by using the Kramers-Kronig analysis (*24*). In Fig. 3B, these curves are transformed into mean refractive indices against optical wavelength. For λ = 550 nm, the mean refractive index is $m = 1.67 - 0.27i$. This conforms to neither the refractive indices of soot carbon [e.g., $m \approx 1.75 - 0.63i$ to $1.95 - 0.79i$ for void-free soot carbon (*4*)] nor the refractive indices for negligibly absorbing organic carbon ($m \approx 1.5 - 0i$). It is also separate from that of the humiclike substances, which have recently been receiving attention as a type of brown carbon because their refractive indices indicate substantially lower optical absorption ($m \approx 1.5 - 0.002i$) (*7*, *29*). Thus, assuming a continuum of carbon species exists (*30*), the carbon spheres described here represent a more strongly absorbing carbon compared with the weakly absorbing, soluble, humiclike substances.

In contrast to laser scattering, our approach determines refractive indices from $\lambda \approx 1200$ nm in the infrared, through the visible ($\lambda \sim 400$ to 700 nm), and into the ultraviolet ($\lambda < 400$ nm), as shown in Fig. 3B. In this figure, it can be seen that optical absorption, represented by k, is low in the infrared and visible red; increases through the visible red, yellow, green, and blue; and peaks in the near-ultraviolet. This increase in absorption with decreasing wavelength indicates that the spheres are constituted by a brown form of carbon, thus confirming their nature as different from that of the black soot carbon. Brown carbon as an aerosol component has recently "come into the forefront of atmospheric research" (*16*). However, its exact nature has remained ill-

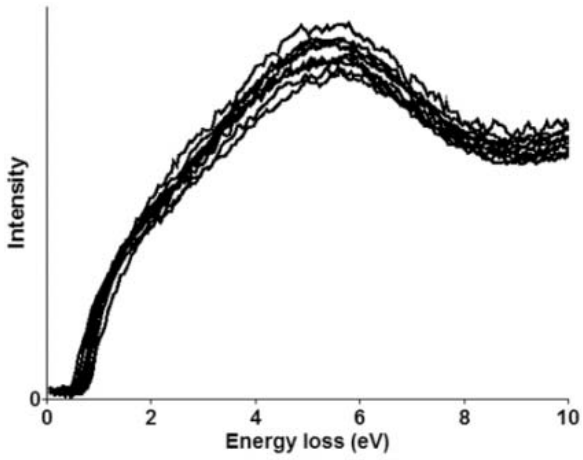


Fig. 2. Experimentally acquired low-loss spectra from typical carbon spheres (deconvoluted spectra to remove the effects of plural scattering and the zero-loss peak) plotted on linear axes. The scattering intensities of the spectra have also been scaled relative to each other in order to aid comparison. The spectra are all similar in form, partly a consequence of the similarity of their material and their optical properties (*24*).

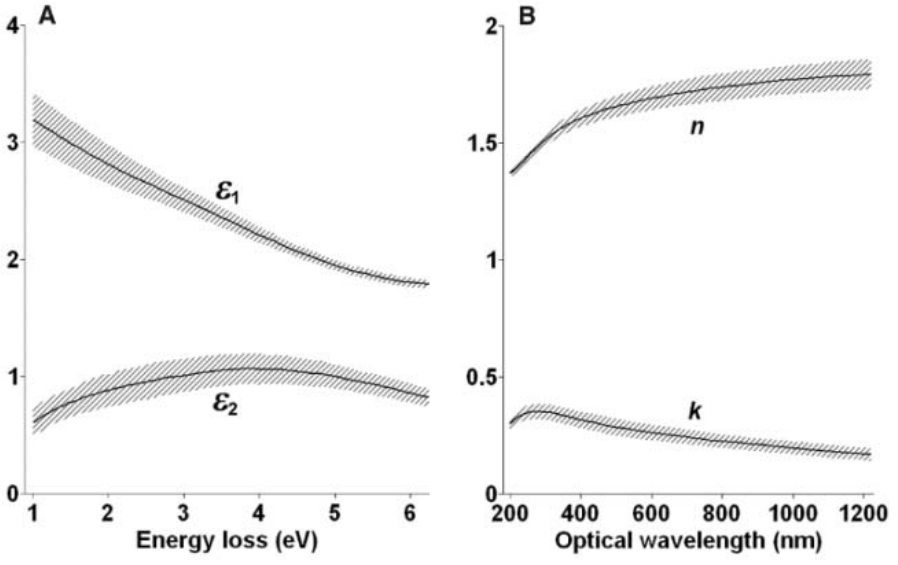


Fig. 3. Mean complex dielectric functions ε_1 and ε_2 against electron energy loss (**A**) and mean refractive indices n and k against optical wavelength (**B**) for 28 analyzed spheres ranging from 130 to 400 nm in diameter (sampled at altitudes of 30 m and 610 m). The shaded bands represent the variance in the mean for the sampling population. The small size of these bands indicates that the spheres have relatively uniform optical properties. k is low in the infrared (>800 nm), increases through the visible, and peaks in the near ultraviolet (<400 nm), thus suggesting that the spheres will appear brown.

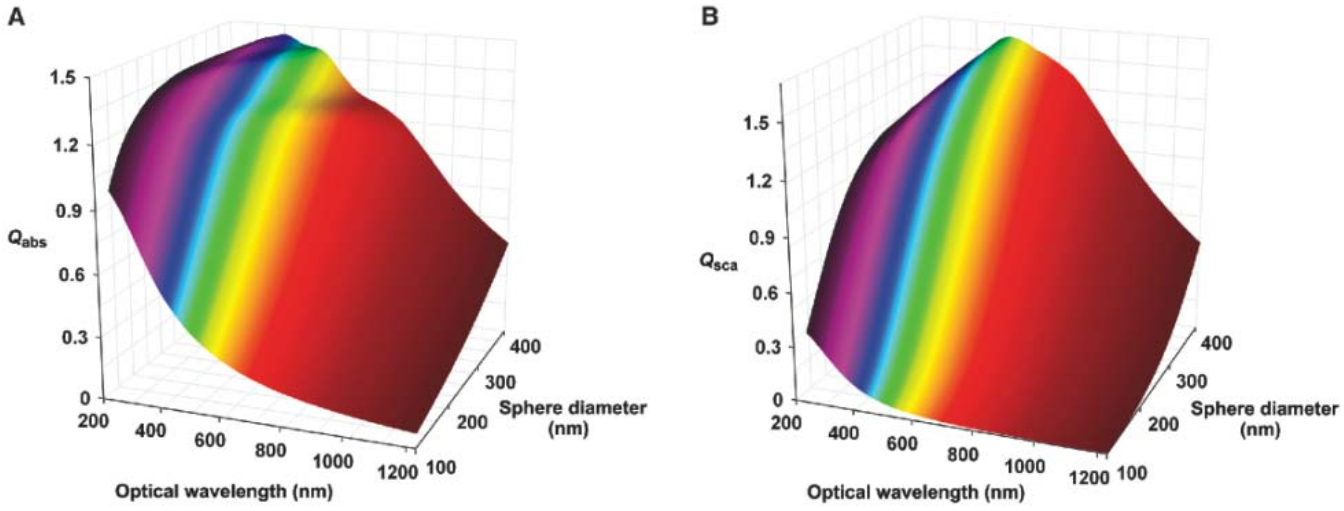


Fig. 4. Diameter and optical wavelength dependence of the light absorbing and scattering efficiencies (*25*), Q_{abs} (**A**) and Q_{sca} (**B**), of the carbon spheres. The calculations of Q_{abs} and Q_{sca} used the mean refractive indices from Fig. 3B. The plots are colored to match the colors of the optical wavelengths; darkening of the red at high wavelengths and of the violet at low wavelengths represent infrared and ultraviolet, respectively.

Table 1. Estimated optical and physical properties for typical carbon sphere and soot carbon particles in the samples studied, calculated for $\lambda = 550$ nm (apart from $\mathring{A}_{abs}$). Assumptions used for determining the soot carbon optical properties are given in the main text.

Property	Carbon sphere	Soot carbon
$N = n - ik$	$1.27 - 0.27i$	$1.95 - 0.79i$
Average spherule diameter (nm)	230	32
Number of spherules	1	55
σ_{abs} ($\times 10^{-14}$ m^2)	4.0	0.8
σ_{sca} ($\times 10^{-14}$ m^2)	3.2	0.2
ω_0	0.44	0.23
ρ (g cm^{-3})	~1.4–1.6	1.8–2.0
MAE (m^2 g^{-1})	~3.6–4.1	4.3–4.8
MSE (m^2 g^{-1})	~3.2–3.7	1.3–1.4
$\mathring{A}_{abs}$	~1.5	~1

defined. Our measurements defined the optical properties of a brown carbon species, an essential step toward modeling and accounting for their effects. For instance, by using Mie theory, in Fig. 4 we model the light absorption and scattering efficiencies (Q_{abs} and Q_{sca}) of the spheres against optical wavelength and sphere diameter, where $Q_{abs} = \sigma_{abs}/\pi r^2$ (absorption cross section, σ_{abs}; sphere radius, r) and analogously $Q_{sca} = \sigma_{sca}/\pi r^2$. As sphere diameter rises, the number of contributing scattering modes increases, increasing absorption and modifying the form of the curves of Q_{abs} and Q_{sca} versus λ. The spectral dependence of absorption is parameterized by the Ångström exponent for absorption ($\mathring{A}_{abs}$), for which $\sigma_{abs} \propto \lambda^{-\mathring{A}_{abs}}$. From Fig. 4, it can be understood that $\mathring{A}_{abs}$ will vary depending on sphere diameter. However, for the measured mean diameter of 230 nm of the spheres analyzed here, $\mathring{A}_{abs} \approx 1.5$ for an optical range of $\lambda = 380$ to 1000 nm, which is consistent with a brown nature and with values for mixed aerosols (*15*).

The EELS carbon K-edge intensity has been used to calculate the density of carbon atoms for the optically analyzed carbon spheres, as described in (*24*). From this, we estimated that their physical densities lie in the range of ~1.4 to 1.6 g cm^{-3}. By using this together with absorption cross sections that we determined using Mie theory, we estimated a mean mass absorption efficiency (MAE) for the spheres to be ~3.6 to 4.1 m^2 g^{-1} for λ = 550 nm. This is slightly lower than ACE-Asia measurements indicating that the MAE for the total light-absorbing carbon content was ~7 ± 2 m^2 g^{-1} (*13*). It is also comparable with estimates of the MAE of soot carbon at $\lambda = 550$ nm (Table 1) (*4*), such that the contribution to atmospheric light absorption by the spheres may not be readily identified or discriminated from that due to soot by the measurement of MAE performed at a single wavelength.

The amorphous carbon spheres make a contribution to direct radiative forcing that is comparable to that of soot. For the samples analyzed here, the average sphere has a diameter of ~230 nm and a refractive index at $\lambda = 550$ nm of $m = 1.67 - 0.27i$. A typical carbon soot particle in the samples consists of an aggregate of ~55 spherules of ~32-nm diameter (derived by an analysis of approximate spherule numbers and sizes in TEM images of 50 randomly chosen soot carbon particles; see Fig. 1B for a typical aggregate). Assuming an upper-bound refractive index of $m = 1.95 - 0.79i$ for the soot (*4*), along with a fractal dimension of 1.8 and a fractal prefactor of 8 (*17*), we calculated absorption cross sections, σ_{abs}, by using Mie theory (*25*) for the spheres and Rayleigh-Debye-Gans theory (*31*) for the soot. The absorption cross section of a typical sphere is almost a factor of 5 greater than that for a typical soot aggregate, that is, $\sigma_{abs} \approx 3.9 \times 10^{-14}$ m^2 (sphere) versus $\sigma_{abs} \approx 0.82 \times 10^{-14}$ m^2 (soot). Other optical and physical properties are compared in Table 1.

Thus, on a particle-by-particle basis, the carbon spheres show light absorption commensurate to, or greater than, that of the soot particles conventionally accepted as being the dominant form of light-absorbing carbon. Our statistical measurements also find that the carbon spheres are very abundant. Specifically, for the nominal particle size range of 0.3 to 2.0 μm studied in these samples, the number ratios of sphere:soot particles is ~1:1 for the specimen sampled at 30 m altitude and ~3:1 for the specimen sampled at 610 m. Moreover, similar carbon spheres have been found abundant or dominant in other regions, such as central Europe and South Africa (*22*), and in urban aerosols in areas such as Phoenix. There may be other ways in which the unique optical properties of the carbon spheres may affect radiative force relative to soot. For instance, owing to their reduced absorption of infrared compared with visible, their role in absorption of reradiated heat will be quite different to that for black soot.

Taking account of both the abundance of spheres and their optical properties, these measurements challenge the usual supposition in radiative forcing calculations of atmospheric carbon being either negligibly absorbing organic carbon or strongly absorbing black soot carbon. For modeling of radiative forcing from aerosols to be accurate, scattering and absorption by brown carbon needs to be explicitly included.

References and Notes

1. V. Ramaswamy *et al.*, in *Climate Change 2001: The Scientific Basis*, J. Houghton *et al.*, Eds. (Cambridge Univ. Press, Cambridge, 2001), pp. 349–416.
2. P. Forster *et al.*, *Changes in Atmospheric Constituents and in Radiative Forcing*, http://ipcc-wg1.ucar.edu/wg1/wg1-report.html (2007).
3. R. C. Millikan, *J. Opt. Soc. Am.* **51**, 698 (1961).
4. T. C. Bond, R. W. Bergstrom, *Aerosol Sci. Technol.* **40**, 27 (2006).
5. E. Dinar *et al.*, *Faraday Discuss.* **137**, 279 (2008).

6. T. W. Kirchstetter, T. Novakov, P. V. Hobbs, *J. Geophys. Res.* **109**, D21208 (2004).
7. A. Hoffer *et al*., *Atmos. Chem. Phys.* **6**, 3563 (2006).
8. H. Horvath, *Atmos. Environ.* **27A**, 293 (1993).
9. J. T. McCartney, S. Ergun, *Fuel* **37**, 272 (1958).
10. A. B. Pluchino, S. S. Goldberg, J. M. Dowling, C. M. Randall, *Appl. Opt.* **19**, 3370 (1980).
11. J. L. Hand *et al*., *J. Geophys. Res.* **110**, D21210 (2005).
12. B. J. Huebert *et al*., *J. Geophys. Res.* **108**, 8633 (2003).
13. A. D. Clarke *et al*., *J. Geophys. Res.* **109**, D15S09 (2004).
14. M. O. Andreae, A. Gelencsér, *Atmos. Chem. Phys.* **6**, 3131 (2006).
15. R. W. Bergstrom *et al*., *Atmos. Chem. Phys.* **7**, 5937 (2007).
16. H. Lukacs *et al*., *J. Geophys. Res.* **112**, D23S18 (2007).
17. G. M. Faeth, Ü. Köylü, *Combust. Sci. Tech.* **108**, 207 (1995).
18. D. M. Roessler, F. R. Faxvog, R. Stevenson, G. W. Smith, in *Particulate Carbon*, D. Siegla, G. Smith, Eds. (Plenum, New York, 1981), pp. 57–89.
19. R. L. Vander Wal, A. J. Tomasek, *Combust. Flame* **136**, 129 (2004).
20. L. H. van Poppel *et al*., *Geophys. Res. Lett.* **32**, L24811 (2005).
21. Y. Chen, N. Shah, A. Braun, F. E. Huggins, G. P. Huffman, *Energy Fuels* **19**, 1644 (2005).
22. M. Pósfai *et al*., *J. Geophys. Res.* **109**, D06213 (2004).
23. J. Hallet, J. G. Hudson, C. F. Rogers, *Aerosol Sci. Technol.* **10**, 70 (1989).
24. Materials and methods are available as supporting material on *Science* Online.
25. C. Bohren, D. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1983).
26. R. H. Ritchie, *Phys. Rev.* **106**, 874 (1957).
27. H. Raether, *Surf. Sci.* **8**, 233 (1967).
28. R. F. Egerton, *Electron Energy-Loss Spectroscopy in the Electron Microscope* (Plenum, New York, ed. 2, 1996).
29. N. Havers, P. Burba, J. Lambert, D. Klockow, *J. Atmos. Chem.* **29**, 45 (1998).
30. H. Sun, L. Bierdermann, T. C. Bond, *Geophys. Res. Lett.* **34**, L17813 (2007).
31. C. M. Sorensen, *Aerosol Sci. Technol.* **35**, 648 (2001).

32. This material is based upon work supported by NASA under grant no. NNG04GB37G issued through the Radiation Sciences Program. Participation in ACE-Asia was funded by NSF grant ATM-0002513. The authors acknowledge support of the National Center for Electron Microscopy (NCEM), Lawrence Berkeley Laboratory, which is supported by the U.S. Department of Energy under contract no DE-AC02-05CH11231. We respectfully acknowledge technical support from the late C. Nelson of the NCEM.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5890/833/DC1
Materials and Methods
Figs. S1 to S3
References

16 January 2008; accepted 2 July 2008
10.1126/science.1155296

A Conserved Mutation in an Ethylene Biosynthesis Enzyme Leads to Andromonoecy in Melons

Adnane Boualem,[1] Mohamed Fergany,[1] Ronan Fernandez,[1] Christelle Troadec,[1] Antoine Martin,[1] Halima Morin,[2] Marie-Agnes Sari,[3] Fabrice Collin,[3] Jonathan M. Flowers,[4] Michel Pitrat,[5] Michael D. Purugganan,[4] Catherine Dogimont,[5] Abdelhafid Bendahmane[1]*

Andromonoecy is a widespread sexual system in angiosperms characterized by plants carrying both male and bisexual flowers. In melon, this sexual form is controlled by the identity of the alleles at the *andromonoecious* (*a*) locus. Cloning of the *a* gene reveals that andromonoecy results from a mutation in the active site of 1-aminocyclopropane-1-carboxylic acid synthase. Expression of the active enzyme inhibits the development of the male organs and is not required for carpel development. A causal single-nucleotide polymorphism associated with andromonoecy was identified, which suggests that the *a* allele has been under recent positive selection and may be linked to the evolution of this sexual system.

Most angiosperms have hermaphroditic flowers containing both male and female organs; nevertheless, sex determination can result in the formation of separate male and female flowers on either the same (monoecy) or different individuals (dioecy). Andromonoecy, where plants carry both male and perfect bisexual flowers, has evolved independently numerous times (*1*) and is found in ~4000 species in 33 angiosperm families (*2*). Several species in the Cucurbitaceae, including cucumber (*Cucumis sativus*) and melon (*Cucumis melo*), have bisexual floral primordia, but often have flowers limited to a single sex. Sex determination occurs by the selective arrest of either the male stamen or female carpel during development (*3*, *4*). In melon, sex determination is governed by the genes *andromonoecious* (*a*) and *gynoecious* (*g*), and the interplay of these two genes results in a range of sexual types (*5*, *6*). Monoecious (*A-G-*) and andromonoecious (*aaG-*) individuals bear male flowers on the main stem and, respectively, female or hermaphrodite flowers on axillary branches, whereas gynoecious (*AAgg*) and hermaphrodite (*aagg*) individuals only bear female and hermaphrodite flowers, respectively (*6*). In addition, cucurbit sex expression patterns can be modified by hormones, such as ethylene, and by environmental factors (*7*, *8*).

In *C. melo*, most plants are monoecious or andromonoecious because of the *a* locus, which has been mapped to a genetic interval of 25.2-centimorgans (*9*). We cloned the *a* locus by constructing high-resolution genetic and physical maps (fig. S1A), used chromosome walking to construct a bacterial artificial chromosome (BAC) contig anchored to the genetic map (fig. S1A), delimited the *a* locus to a single BAC clone, and revealed seven candidate genes within 107 kilobase pairs (kbp) (fig. S1B) (*10*). The *a* locus flanking marker sequences L41 and R5 were used to identify a 14-kbp region containing a gene encoding for a 1-aminocyclopropane-1-carboxylic acid synthase (ACS), designated *CmACS-7* on the basis of homology to the *Arabidopsis ACS-7* gene (At4g26200). ACS is a pyridoxal 5′-phosphate (PLP)-dependent enzyme that catalyzes the first committed, and generally rate-limiting, step in the production of 1-aminocyclopropane-1-carboxylic acid (ACC) from *S*-adenosylmethionine (SAM) in ethylene biosynthesis in higher plants (*11*). Ethylene is then made from ACC by the enzyme ACC oxidase.

A TILLING (targeting induced local lesions in genomes) approach confirmed the role of *CmACS-7* in sex determination (*12*). We identified six mutations in the full *CmACS-7* gene; four silent or intronic and two that led to missense mutations at G19E and D376N (*13*). The G19E change occurs in a highly conserved amino acid position and may affect the function of the protein (*14*), whereas the D376N modification affects a non-conserved amino acid position (Fig. 1A). Backcrosses to the wild type showed that, for more than 100 F_2 plants for each cross, that the D376N mutation, as well as the silent and intronic mutations, had no sexual phenotype (Fig. 1B and fig. S2). In contrast, plants homozygous for the G19E mutation were andromonoecious (Fig. 1B and fig. S2). On the basis of these data, we concluded that *CmACS-7* is the *andromonoecious* gene.

A 14-kbp genomic sequence of melon accessions PI124112 and Védrantais was used to map single-nucleotide polymorphisms (SNPs) linked to *CmACS-7*. The minimal interval between allelic variants was localized to a 2438-bp DNA fragment containing 560 bp of the proximal promoter, as well as exons 1 and 2 and a part of exon 3 (fig. S1C). We identified no differences linked to sex phenotype in the proximal promoter sequence and quantitative reverse transcription polymerase chain reaction (RT-PCR) detected no significant

[1]INRA (Institut National de la Recherche Agronomique)–CNRS, UMR1165, Unité de Recherche en Génomique Végétale, 2 rue Gaston Crémieux, F-91057 Evry, France. [2]INRA, UR501, Laboratoire de Biologie Cellulaire, Institut Jean Pierre Bourgin, F-78026 Versailles, France. [3]CNRS, UMR 8601, Laboratoire de Chimie et Biochimie Pharmacologiques et Toxicologiques, Université René Descartes, 45 Rue des Saint-Pères, F-75270 Paris Cedex 06, France. [4]Department of Biology, Center for Genomics and Systems Biology, 100 Washington Square East, New York University, New York, NY 10003, USA. [5]INRA, UR 1052, Unité de Génétique et d'Amélioration des Fruits et Légumes, BP 94, F-84143 Montfavet, France.

*To whom correspondence should be addressed. E-mail: bendahm@evry.inra.fr

difference in the expression level between *aa* and *AA* individuals (Fig. 2A). However, we identified a single missense mutation, A57V, in the CmACS-7 protein (Fig. 1A and fig. S1, C to E). A^{57} is conserved across the ACS protein in seed plants (*15*) (Fig. 1A), and crystallographic studies have determined that the backbone nitrogen of A^{57} forms hydrogen bonds with the α-carboxylate group of the enzyme substrate SAM (*16*), which suggests that it may affect protein function (fig. S3).

To determine whether the andromonoecious phenotype is due to a loss of CmACS-7 enzymatic activity, we assayed ACS activity in vitro by monitoring 5′-methylthioadenosine (MTA) formation at different PLP concentrations. In the presence of high concentrations of PLP, both enzymes showed similar activity (table S1, fig. S4) but at a physiological PLP concentration [~1 μM in the flowers (fig. S5)], the V^{57} ACS isoform activity is undetectable, whereas the A^{57} form retains 50% of its maximum activity (fig. S4). We found that, to reach maximum activity, the V^{57} form needs PLP in amounts greater than 100 times that found naturally (Fig. 1C). These data are consistent with the binding constant (K_d) observed upon PLP binding to the enzyme (table S1). Because the interaction between A^{57} and PLP has not previously been demonstrated, the observed decrease in binding affinity may be due to an improper orientation of SAM in the active site—which may be required for the formation of external aldimine between PLP and SAM (*16*) (fig. S3). We also saw that ACS activity was undetectable in the G19E mutant (Fig. 1C). Thus, on the basis of the loss of the enzymatic activity of the forms containing the natural A57V and the induced G19E mutation, we concluded that an active CmACS-7 enzyme is required for the development of female flowers in monoecious lines, whereas a reduction of enzymatic activity results in hermaphrodite flowers in andromonoecious lines.

In melon female flowers, stamen arrest occurs at stage 6 just after the elaboration of carpel primordia (*17*). Quantitative RT-PCR (Fig. 2A) and in situ hybridization demonstrate that *CmACS-7* mRNA is mainly expressed in female and hermaphrodite flowers at stage 4, when flowers are not morphologically distinguishable (Fig. 2, C and D), and at later stages (Fig. 2, F and G). The early accumulation of *CmACS-7* mRNA was strongly localized in the carpel primordia (Fig. 2, C, D, F, and G). In the male flowers of the PI124112 and Védrantais lines, *CmACS-7* mRNA was not detected at any developmental stages (Fig. 2, B and E). Because *CmACS-7* expression level and pattern were not different between female and hermaphrodite flowers (Fig. 2) and because the loss of CmACS-7 activity accounts for the functional variation, we concluded that CmACS-7–mediated ethylene production in the carpel primordia affects the development of the stamina in female flowers but is not required for carpel development.

To assess the level of polymorphism at *CmACS-7*, we sequenced the entire coding region in a panel of 106 *C. melo* accessions, including 37 monoecious (*AAGG*), 3 gynoecious (*AAgg*), 65 andromonoecious (*aaGG*), and 1 hermaphrodite (*aagg*) individuals (table S2). We observed 21 polymorphisms in the noncoding regions and 9 in the coding sequence. Of the nine coding polymorphisms, two produced amino acid changes, the A57V substitution we identified as responsible for the andromonoecious phenotype and a T230S modification that was observed in only one monoecious accession in a region that is variable among seed plants. The A57V substitution in *CmACS-7* produces a diagnostic loss of an Alu 1 restriction site (fig. S6). We analyzed a germplasm collection of 497 accessions of *C. melo* from different parts of the world with an Alu 1 cleaved amplified polymorphic sequence (CAPS) marker (table S2). All monoecious ($n = 146$) and gynoecious ($n = 3$) accessions identified as harboring the dominant allele *A* contained the A^{57}

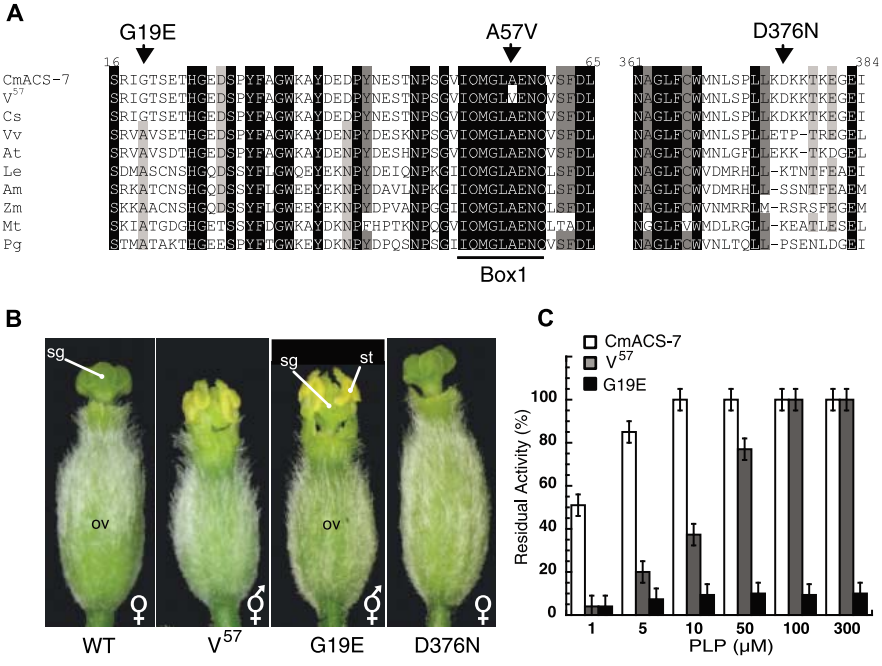


Fig. 1. (**A**) Amino acid alignments of the CmACS-7 and V^{57} isoform with homologous proteins from *Cs* (*Cucumis sativus*, BAF79596), *Vv* (*Vitis vinifera*, CAN66901), *At* (*Arabidopsis thaliana*, AF322390), *Le* (*Lycopersicon esculentum*, AF179247), *Am* (*Antirrhinum majus*, AAC70353), *Zm* (*Zea mays*, AAR25560), *Mt* (*Medicago truncatula*, AAL35745), and *Pg* (*Picea glauca*, ABM60747). Numbers above the alignment indicate the amino acid positions along the CmACS-7 protein. Box 1 indicates a conserved domain in ACS. A57V sequence variation and G19E and D376N EMS-induced mutations are shown above the alignment. (**B**) Flower types from monoecious (WT), andromonoecious (V^{57}), and the G19E and D376N mutants. Male flowers are identical in all the melon types (see SOM). Labels: ov, ovary; sg, stigma; and st, stamen. (**C**) Effect of PLP concentration in the presence of 60 μM SAM on the ACS enzymatic activity of CmACS-7, V^{57}, and G19E protein forms. Residual activities (%) corresponds to the specific activity measured divided by the specific activity measured for CmACS-7 enzyme at 300 μM PLP.

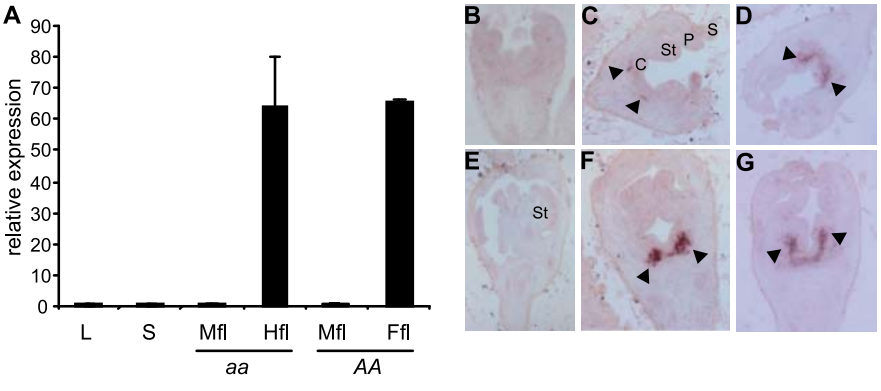


Fig. 2. (**A**) Quantitative RT-PCR of *CmACS-7* in different organs, means ± SD of three biological replicates. L, leaf; S, stem; Mfl, male flower; Hfl, hermaphrodite flower; and Ffl, female flower. Genotypes are indicated in italics below the graph. (**B** to **G**) *CmACS-7* in situ expression at stages 4 (B, C, D) and 7 (E, F, G) of flower development. Male flowers (B and E), hermaphroditic flowers (C and F), and female flowers (D and G). C, carpel; St, stamen; P, petal; and S, sepal.

diagnostic Alu 1 site; all andromonoecious ($n = 347$) and hermaphrodite individuals containing the recessive allele *a* showed results consistent with andromonoecious sex determination via the V^{57} sequence.

To define haplotypes across the entire *CmACS-7* genomic sequence, we used all identified SNPs between the PI124112 and Védrantais lines. Out of the 106 accessions sequenced for *CmACS-7*, a single haplotype, H12, occurred in all the hermaphrodite and andromonoecious accessions, whereas the monoecious and gynoecious accessions contained 11 unique haplotypes (H1 to H11) (Fig. 3). We also compared 1.6 kbp of the promoter region in a subset of 35 melon accessions, but found no differences between hermaphrodite and andromonoecious accessions. In fact, the promoters of the monoecious lines were highly variable. Interestingly, in one monoecious accession, Cantaloup d'Alger, the 1.6-kbp promoter was identical to that of the andromonoecious accessions (fig. S7), which supports the finding that the andromonoecious *CmACS-7* allele is in the coding region.

A maximum parsimony phylogenetic tree created on the basis of the *CmACS-7* gene sequence among accessions indicates that the *a* allele is monophyletic with a single origin derived from the *A* allele (fig. S8). The nucleotide diversity level (π) for this gene in monoecious accessions was 0.0031 (Table 1 and fig. S9A). In contrast, we found no variation in the andromonoecious accessions, consistent with a selective sweep at the *CmACS-7* gene. A haplotype test (*18*) performed under the conservative assumption of no recombination found that the single haplotype within 24 andromonoecious accessions rejected the hypothesis that this site is evolving under neutral evolution ($P < 0.021$). The observed lack of variation at *CmACS-7* was not shared by seven other genes in the melon genome, which suggests that this pattern is not the result of demographic forces. We also conducted a maximum likelihood multilocus HKA test (*19*, *20*) across all six reference genes for which we could obtain the sequence of *C. sativus* ortholog and *CmACS-7*. In this test, a model with selection at *CmACS-7* is significant for andromonoecious accessions compared to a model of neutral evolution across all loci [$P < 0.001$, likelihood ratio test (LRT) = 11.91, df = 1). Finally, the fixation index F_{ST}, a standardized measure of the genetic variance among populations, between andromonoecious and monoecious accessions was highest at *CmACS-7* (mean $F_{ST} = 0.49$) compared with the other genes in the genome (mean $F_{ST} = 0.15$) (Table 1 and fig. S9B). These results are consistent with recent positive selection at the *andromonoecious* allele of the *CmACS-7* gene in *C. melo*.

Linkage between plant hormone synthesis and sex determination also occur in *Mercurialis annua*, where a sex determination gene correlates with the level of zeatin (*21*) and maize, where several mutations affecting sex expression have been identified (*22*). The levels and patterns of polymorphism identified at *CmACS-7* gene are consistent with a recent origin of the andromonoecious allele in this species. These results also suggest that there has been selection for andromonoecy in *C. melo*, which may be due to either selection for resource allocation (*1*) (as males are hypothesized to be less resource intensive than females) or due to the pollen donation hypotheses (*23*) (which suggests that there is a positive correlation between the number of male flowers and male fitness). Both possibilities may be of particular relevance given that *C. melo* is a cultivated species, although greater fruit production in either natural or cultivated species would provide clear selective advantages. Nevertheless, evolutionary ecology theory and experimentation are required to model and assess the selective advantages of andromonoecy in *Cucumis* species, as well as in other plants.

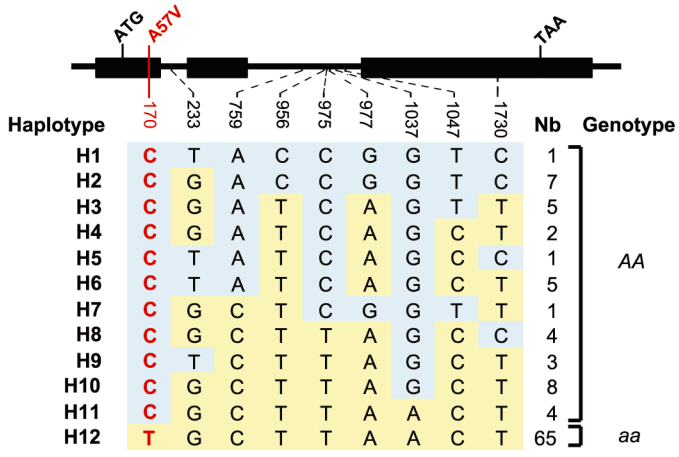


Fig. 3. Haplotypes identified at *CmACS-7*. Each SNP was positioned relative to the first nucleotide of the start codon in PI124112. Nb indicates the number of accessions in each haplotype. Genotype indicates whether the accessions harbor the dominant *A* or the recessive *a* alleles. The SNP at A57V linked with andromonoecy is colored red.

Table 1. Molecular diversity in *CmACS-7* and other melon genes. Nucleotide diversity level (π), Tajima's D, and population differentiation (F_{ST}) levels are indicated. Andro, andromonoecious; mono, monoecious; *Adh1*, alcohol dehydrogenase; *Per-like*, ascorbate peroxidase–like; *Cat2*, catalase 2; *mitoATPg*, mitochondrial ATP synthase g subunit. N.A., not applicable.

Gene	π			Tajima's D			F_{ST}
	Total	Andro	Mono	Total	Andro	Mono	
CmACS-7	0.0023	0.0000	0.0031	−0.9403	n.a.	−0.5937	0.49
Adh-1	0.0027	0.0018	0.0031	−0.1455	−0.4736	−0.281	0.18
Per-like	0.0010	0.0005	0.0014	−0.4764	1.5961	0.0446	0.04
mitoATPg	0.0022	0.0010	0.0027	0.1929	−1.2375	0.3785	0.27
Calmodulin	0.0061	0.0046	0.0069	1.6144	0.5731	1.5624	0.11
Cat2	0.0026	0.0021	0.0027	0.6451	2.3719	0.2689	0.12
Expressed gene	0.0007	0.0002	0.0010	−0.7633	−0.2484	−0.6787	0.32
Shaggy-like	0.0006	0.0003	0.0009	−1.6159	−0.7683	−1.5488	0.00

References and Notes

1. J. S. Miller, P. K. Diggle, *Am. J. Bot.* **90**, 707 (2003).
2. E. Yampolsky, H. Yampolsky, *Bibl. Genet.* **3**, 1 (1922).
3. S. Malepszy, K. Niemirowicz-Szczytt, *Plant Sci.* **80**, 39 (1991).
4. M. M. Kater, J. Franken, K. J. Carney, L. Colombo, G. C. Angenent, *Plant Cell* **13**, 481 (2001).
5. C. F. Poole, P. C. Grimball, *J. Hered.* **30**, 21 (1939).
6. D. Kenigsbuch, Y. Cohen, *Genome* **33**, 317 (1990).
7. R. E. Byers, L. R. Baker, H. M. Sell, R. C. Herner, D. R. Dilley, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 717 (1972).
8. T. Yin, J. A. Quinn, *Am. J. Bot.* **82**, 1537 (1995).
9. C. Périn *et al.*, *Theor. Appl. Genet.* **104**, 1017 (2002).
10. Materials and methods are available as supporting material on *Science* Online.
11. P. W. Oeller, M. W. Lu, L. P. Taylor, D. A. Pike, A. Theologis, *Science* **254**, 437 (1991).
12. C. M. McCallum, L. Comai, E. A. Greene, S. Henikoff, *Nat. Biotechnol.* **18**, 455 (2000).
13. G19E describes a mutant in which glycine is replaced by glutamic acid at position 19. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
14. P. C. Ng, S. Henikoff, *Nucleic Acids Res.* **31**, 3812 (2003).
15. S. G. Ralph, J. W. Hudgins, S. Jancsik, V. R. Franceschi, J. Bohlmann, *Plant Physiol.* **143**, 410 (2007).
16. Q. Huai *et al.*, *J. Biol. Chem.* **276**, 38210 (2001).
17. S. L. Bai *et al.*, *Planta* **220**, 230 (2004).
18. R. R. Hudson, K. Bailey, D. Skarecky, J. Kwiatowski, F. J. Ayala, *Genetics* **136**, 1329 (1994).
19. R. R. Hudson, M. Kreitman, M. Aguade, *Genetics* **116**, 153 (1987).
20. S. I. Wright, B. Charlesworth, *Genetics* **168**, 1071 (2004).
21. B. Durand, R. Durand, *Plant Sci.* **80**, 49 (1991).
22. R. J. Bensen *et al.*, *Plant Cell* **7**, 75 (1995).
23. M. Vallejo-Marin, M. D. Rausher, *Evol. Int. J. Org. Evol.* **61**, 404 (2007).

24. We thank Laboratoire ASL, GAUTIER Semences, and TAKII Research France for the mutant populations; P. Audigier for plant care; and C. Clepet, F. Frugier, and M. Crespi for discussions and comments on the manuscript. Supported by GAP (Plant Breeding and Genetics) department in INRA. Sequences have been deposited in GenBank, accession numbers EU791279, EU791280, and EU798301 to EUT98691.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5890/836/DC1
Materials and Methods
SOM Text
Figs. S1 to S9
Tables S1 to S3
References

11 April 2008; accepted 12 June 2008
10.1126/science.1159023

Human *CHN1* Mutations Hyperactivate α2-Chimaerin and Cause Duane's Retraction Syndrome

Noriko Miyake,[1,2] John Chilton,[3]* Maria Psatha,[4]* Long Cheng,[1,2] Caroline Andrews,[1,2,5] Wai-Man Chan,[1] Krystal Law,[1]† Moira Crosier,[6] Susan Lindsay,[6] Michelle Cheung,[4] James Allen,[3] Nick J. Gutowski,[7,8] Sian Ellard,[8,9] Elizabeth Young,[8] Alessandro Iannaccone,[10] Binoy Appukuttan,[11] J. Timothy Stout,[11] Stephen Christiansen,[12] Maria Laura Ciccarelli,[13] Alfonso Baldi,[14] Mara Campioni,[14] Juan C. Zenteno,[15] Dominic Davenport,[4] Laura E. Mariani,[5] Mustafa Sahin,[2,5] Sarah Guthrie,[4] Elizabeth C. Engle[1,2,5,16,17]‡

Duane's retraction syndrome (DRS) is a complex congenital eye movement disorder caused by aberrant innervation of the extraocular muscles by axons of brainstem motor neurons. Studying families with a variant form of the disorder (DURS2-DRS), we have identified causative heterozygous missense mutations in *CHN1*, a gene on chromosome 2q31 that encodes α2-chimaerin, a Rac guanosine triphosphatase–activating protein (RacGAP) signaling protein previously implicated in the pathfinding of corticospinal axons in mice. We found that these are gain-of-function mutations that increase α2-chimaerin RacGAP activity in vitro. Several of the mutations appeared to enhance α2-chimaerin translocation to the cell membrane or enhance its ability to self-associate. Expression of mutant α2-chimaerin constructs in chick embryos resulted in failure of oculomotor axons to innervate their target extraocular muscles. We conclude that α2-chimaerin has a critical developmental function in ocular motor axon pathfinding.

Ocular motility and binocular vision depend on the precise innervation of six extraocular muscles by the oculomotor, trochlear, and abducens cranial motor neurons (fig. S1A) (*1*). Disruptions in these developmental processes can cause complex congenital eye movement disorders (*2*, *3*), the most common of which is Duane's retraction syndrome (DRS) with an incidence in the general population of about 0.1%. Individuals with DRS have restricted abduction and in some cases adduction of their eyes, with retraction of the globe upon attempted adduction. Postmortem studies of sporadic DRS reveal absence of the abducens motor neurons and cranial nerve, with anomalous innervation of its target, the lateral rectus muscle, by a branch of the oculomotor nerve (fig. S1B) (*4*, *5*).

Four pedigrees (IJ, UA, JH, and FY; fig. S1D) segregating a DRS variant as a dominant trait are reported to map to the DURS2 locus on chromosome 2q31 (*6*–*8*). Examinations of affected family members established that although some have a phenotype indistinguishable from sporadic DRS, overall they have a higher incidence of bilateral involvement and of vertical movement abnormalities (Fig. 1A) (*8*–*10*). Consistent with these clinical findings, our magnetic resonance (MR) imaging of members of pedigrees FY and JH revealed that, in addition to absent or hypoplastic abducens nerves and aberrant lateral rectus innervation by the oculomotor nerve, some individuals had hypoplastic oculomotor nerves and small oculomotor-innervated muscles (*10*). Thus, mutations in the DURS2 gene appear to affect

[1]Department of Medicine (Genetics), Children's Hospital Boston, Boston, MA 02115, USA. [2]Harvard Medical School, Boston, MA 02115, USA. [3]Institute of Biomedical and Clinical Science, Peninsula Medical School, Research Way, Plymouth PL6 8BU, UK. [4]MRC Centre for Developmental Neurobiology, King's College, Guy's Campus, London SE1 1UL, UK. [5]Department of Neurology, Children's Hospital Boston, Boston, MA 02115, USA. [6]MRC-Wellcome Trust Human Developmental Biology Resource (Newcastle), Institute of Human Genetics, Newcastle University, International Centre for Life, Newcastle upon Tyne NE1 3BZ, UK. [7]Department of Neurology, Royal Devon and Exeter Hospital, Barrack Road, Exeter, Devon EX2 5DW, UK. [8]Peninsula Medical School, Barrack Road, Exeter EX2 5DW, UK. [9]Department of Molecular Genetics, Royal Devon and Exeter Hospital, Barrack Road, Exeter, Devon EX2 5DW, UK. [10]University of Tennessee Health Science Center, Hamilton Eye Institute, 930 Madison Avenue, Suite 731, Memphis, TN 38163, USA. [11]Casey Eye Institute, Oregon Health and Science University, 3375 SW Terwilliger Boulevard, Portland, OR 97239, USA. [12]Department of Ophthalmology, University of Minnesota, MMC 493, 420 Delaware Street SE, Minneapolis, MN 55455, USA. [13]Fatebenefratelli Hospital, Division of Ophthalmology, Isola Tiberina, Rome 00100, Italy. [14]Department of Biochemistry "F. Cedrangolo," Section of Pathologic Anatomy, Second University of Naples, Naples 80138, Italy. [15]Department of Genetics and Research Unit, Institute of Ophthalmology "Conde de Valenciana," Mexico City 06700, Mexico. [16]Department of Ophthalmology, Children's Hospital Boston, Boston, MA 02115, USA. [17]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA.

*These authors contributed equally to this work.
†Deceased.
‡To whom correspondence should be addressed. E-mail: elizabeth.engle@childrens.harvard.edu

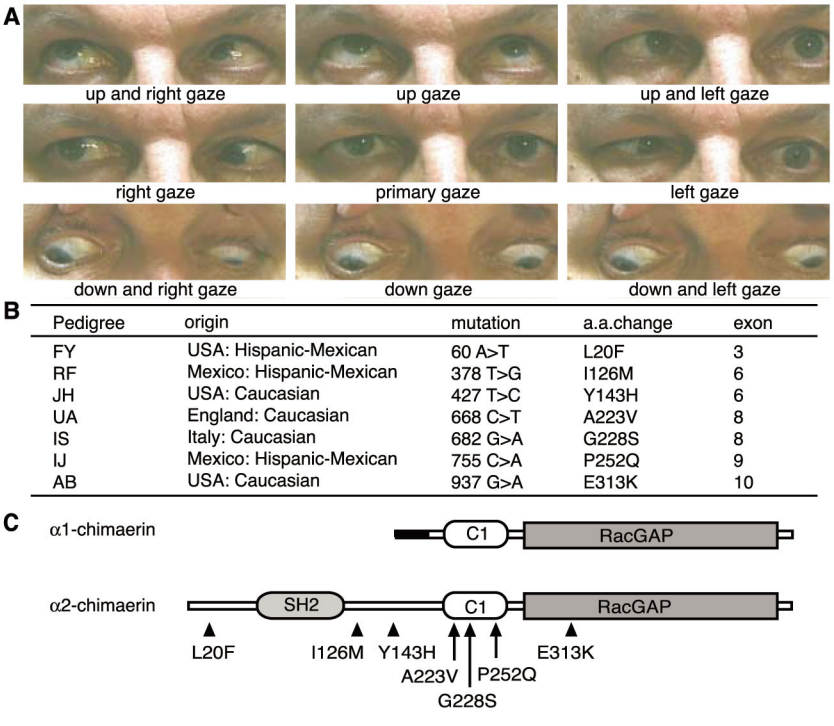


Fig. 1. Duane's retraction syndrome (DRS) and corresponding *CHN1* mutations. (**A**) Affected member of pedigree JH with limited outward gaze (abduction) and narrowing of the palpebral fissure on attempted inward gaze (adduction) most obvious on leftward gaze. He also has bilateral exotropia on downgaze. (**B**) The seven DURS2-DRS pedigrees and corresponding heterozygous *CHN1* mutations. (**C**) Schematic representation of α1-chimaerin (top, 334 amino acids) and α2-chimaerin (bottom, 459 amino acids) protein. The isoforms contain identical C1 and RacGAP domains; only α2-chimaerin contains an SH2 domain. Mutations alter residues unique to α2-chimaerin or common to both proteins, as indicated by the arrows. No mutations were found in the α1-chimaerin N-terminal sequence (highlighted in black).

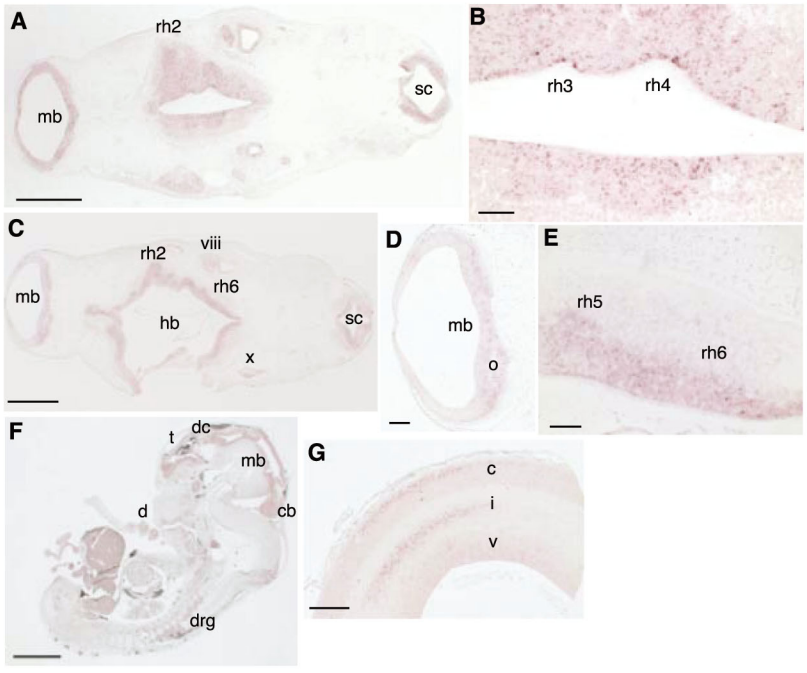


Fig. 2. Human developmental expression profile of α2-chimaerin mRNA by in situ hybridization. (**A**) Transverse section of CS15 human embryo showing α2-chimaerin mRNA expression (purple deposit) in midbrain, hindbrain (rhombomere 2 indicated), and spinal cord. (**B**) Higher magnification of (A) showing expression in the ventricular layer of rhombomeres 3 and 4. (**C**) At CS16, expression is also seen in midbrain, hindbrain, spinal cord, and vestibulocochlear (viii) and vagus (x) nuclei. Higher magnifications of (C) show (**D**) expression in developing oculomotor neurons and (**E**) in neurons of rhombomeres 5 (developing abducens neurons) and 6. In CS19 sagittal section (**F**), expression has declined in basal midbrain and hindbrain and is now found in dorsal root ganglia, cerebellum, diencephalon, and telencephalon. At later stages (**G**), expression is located in specific regions of the cortical plate and the intermediate and ventricular zones of the forebrain (11 weeks post-ovulation). No signal was detected in corresponding sections hybridized with sense probe (fig. S4). Abbreviations: mb, midbrain; hb, hindbrain; rh, rhombomere; sc, spinal cord; o, oculomotor nucleus; drg, dorsal root ganglia; d, digit; cb, cerebellum; dc, diencephalon; t, telencephalon; c, cortical plate; i and v, intermediate and ventricular zones of the forebrain. Scale bars, 1000 μm [(A) and (C)], 100 μm [(B) and (E)], 200 μm (D), 2000 μm (F), 500 μm (G).

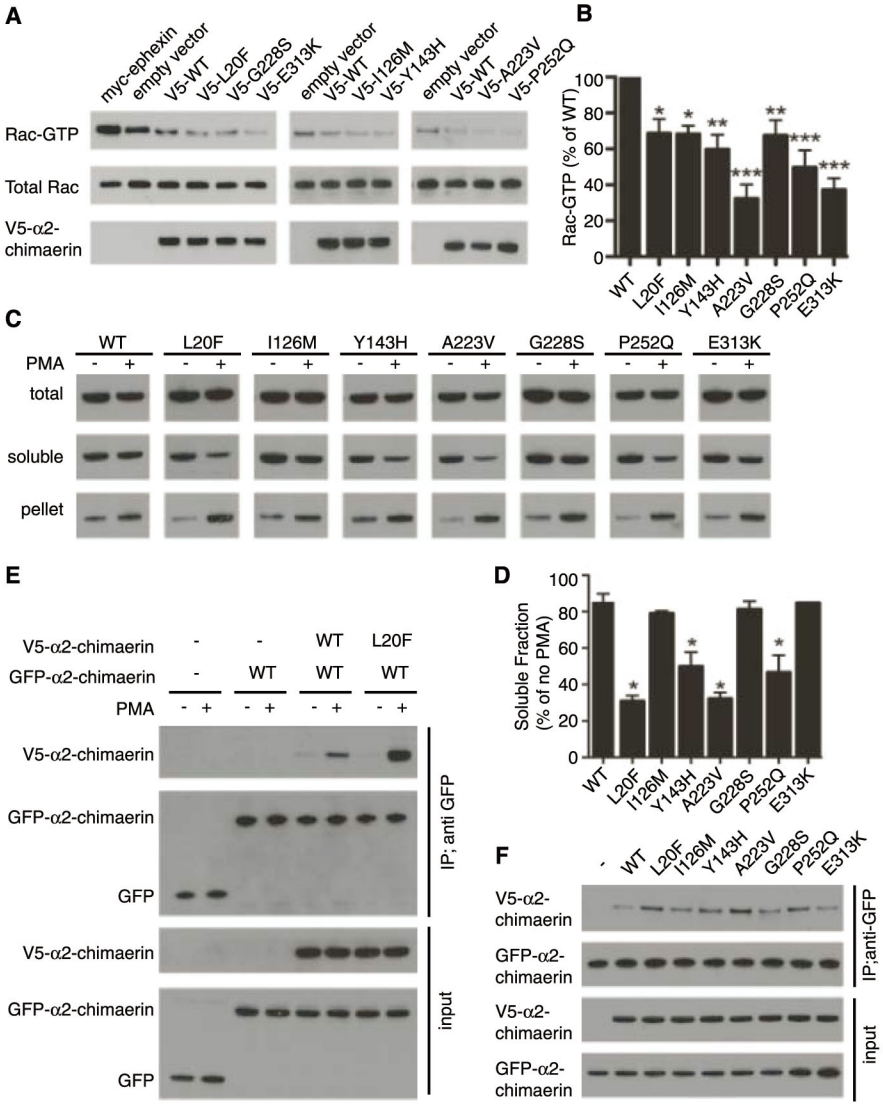


Fig. 3. DURS2-DRS mutations enhance α2-chimaerin function in vitro. (**A**) Rac-GTP levels were measured in HEK 293T cells transfected with plasmids encoding myc-ephexin, epitope-tagged V5–empty vector, V5–α2-chimaerin wild-type, or V5–α2-chimaerin mutants. Rac-GTP levels are reduced by overexpression of wild-type α2-chimaerin relative to empty vector, and are further reduced in cells expressing each mutant, but are elevated with overexpression of a guanine nucleotide exchange factor, myc-ephexin (*27*). (**B**) Densitometric analysis of Rac-GTP levels normalized to total Rac and V5–α2-chimaerin levels. Values are expressed as percent of wild-type α2-chimaerin (mean ± SEM, n = 6 to 10). The difference between the reduction of Rac-GTP levels for each mutant compared to wild-type α2-chimaerin is significant by one-way analysis of variance (ANOVA) with Dunnett's adjustment (F = 9.89, *P < 0.03, **P < 0.005, ***P < 0.0001). (**C**) α2-Chimaerin translocation examined by immunoblots of total, soluble, and pellet fraction of wild-type and mutant α2-chimaerin with or without 10 μM PMA stimulation. (**D**) Graphical representation of translocation after PMA treatment, expressed as percent of α2-chimaerin remaining in the soluble fraction (mean ± SEM, n = 3). Enhanced translocation compared to wild-type is significant for L20F, Y143H, A223V, and P252Q by one-way ANOVA with Dunnett's adjustment (F = 21.00, *P < 0.0001). (**E**) GFP–α2-chimaerin immunoprecipitates with V5–wild type or V5-L20F α2-chimaerin in the presence of PMA, and minimally in its absence. (**F**) In the presence of PMA, immunoprecipitation of wild-type α2-chimaerin is enhanced by all mutant α2-chimaerins relative to the wild type except G228S and E313K, which were equivalent to wild-type α2-chimaerin. Results were consistent over at least four independent experiments (see also fig. S6, F and G).

primary development of the abducens and, to a lesser degree, the oculomotor nerve (fig. S1C).

To identify the DURS2 gene, we further analyzed the recombination events that defined the published DURS2 critical region (*6*, *7*), reducing it from 9.9 to 4.6 Mb (fig. S2, A and B), and then sequenced 22 positional candidate genes (fig. S2B) in a proband from each of the four published pedigrees. We identified in each a unique heterozygous missense change in *CHN1*, which encodes two Rac-specific guanosine triphosphatase (GTPase)–activating α-chimaerin isoforms. We then screened 16 smaller pedigrees that segregated DRS in a dominant fashion, and identified three additional heterozygous *CHN1* missense changes in pedigrees RF, IS, and AB (Fig. 1B and figs. S1E and S2C). All seven nucleotide substitutions cosegregated with the affected haplotypes, and none were present in online single-nucleotide polymorphism databases or on 788 control chromosomes. Five of the substitutions are predicted to result in nonconservative amino acid substitutions [Leu^{20} → Phe (L20F), Tyr^{143} → His (Y143H), Gly^{228} → Ser (G228S), Pro^{252} → Gln (P252Q), and Glu^{313} → Lys (E313K)] and two in conservative amino acid substitutions [Ile^{126} → Met (I126M) and Ala^{223} → Val (A223V)] (Fig. 1B). All are predicted to alter amino acids that are conserved in eight different species (fig. S2D).

The Rho family member Rac is a GTPase that is active when GTP-bound; it serves as a regulator of downstream intracellular signaling cascades controlling cytoskeleton dynamics, including the growth and development of dendrites and axons. Rac is inactivated by 12 Rac GTPase-activating proteins (GAPs) in the mammalian genome (*11*), including α1- and α2-chimaerin (encoded by *CHN1*) and paralogs β1- and β2-chimaerin (encoded by *CHN2*). In the rodent brain, α2-chimaerin has been shown to serve as an effector for axon guidance (*12*–*16*), whereas α1-chimaerin appears to play a later role in dendritic pruning (*17*, *18*).

CHN1 is alternatively spliced, and the α1-chimaerin promoter lies in intronic sequence upstream of α2-chimaerin exon 7 (*19*). Thus, the two isoforms share a RacGAP domain that interacts with and down-regulates Rac activity, as well as a C1 domain that binds to diacylglycerol (DAG), a membrane-associated phorbol ester signaling lipid. Only α2-chimaerin contains an N-terminal Src homology 2 (SH2) domain (*20*, *21*). Three DURS2 mutations alter amino acids unique to α2-chimaerin, whereas four alter residues shared by α1- and α2-chimaerin (Fig. 1C and table S1). Because we cannot distinguish between the two groups clinically, the DURS2 phenotype most likely results from altered α2-chimaerin function.

In situ studies in rats (*20*, *21*) revealed widespread embryonic neuronal expression of α2-chimaerin mRNA. Expression in the caudal brainstem and cephalic flexure peaked at embryonic day (E) 12.5, whereas mouse embryonic expression peaked overall at E10.5 (fig. S3, A and B), both consistent with expression of α2-chimaerin in developing ocular motor neurons. We found similar widespread expression of α2-chimaerin mRNA during human development, strongest at Carnegie stage (CS) 15 and CS16 in the midbrain and hindbrain (Fig. 2, fig. S3, C to E, and fig. S4). Therefore, although expressed in developing ocular motor neurons, the expression pattern alone does not account for the striking restriction of the DURS2 phenotype.

All seven amino acids altered by DURS2 mutations are conserved in α2-chimaerin's paralog,

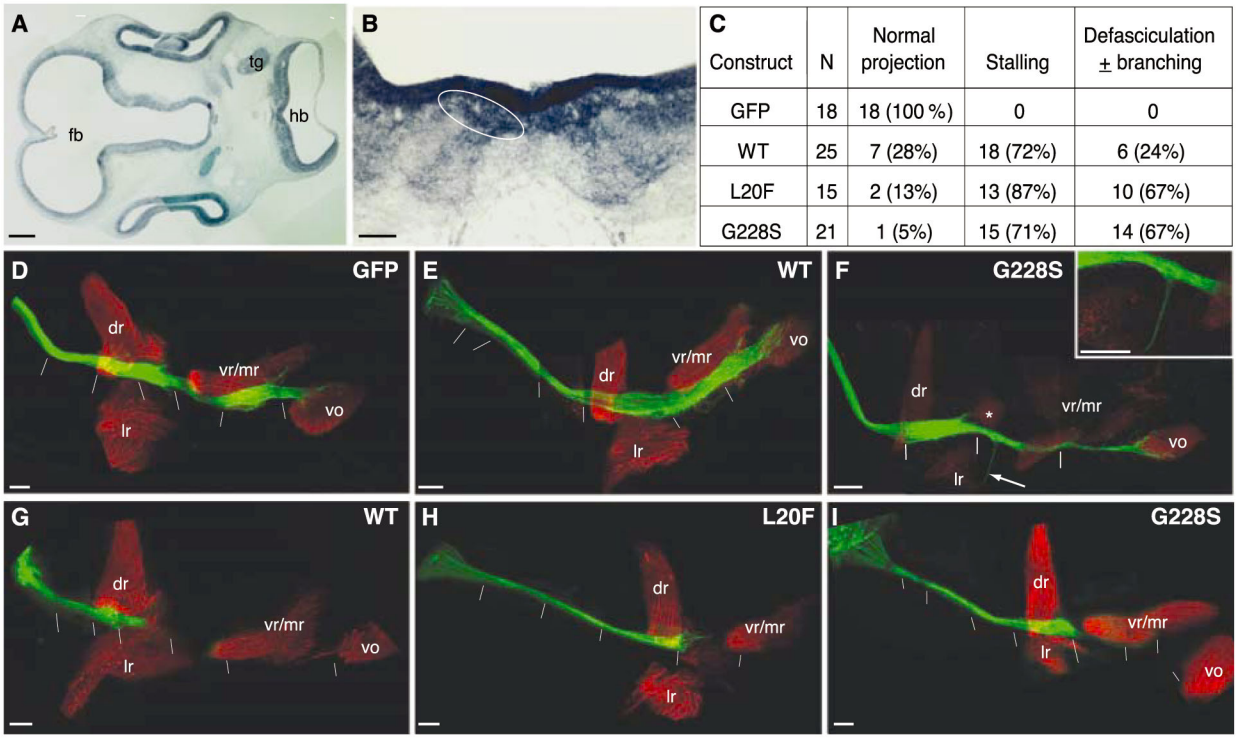


Fig. 4. α2-Chimaerin overexpression results in stalling of developing chick oculomotor nerves. (**A**) Transverse section through E4 whole chick embryo, showing wide neuroepithelial expression of α2-chimaerin mRNA including the hindbrain (hb), forebrain (fb), and trigeminal ganglion (tg). (**B**) Transverse section through E5–6 chick midbrain, showing α2-chimaerin mRNA expression in the oculomotor nuclei (left nucleus circled in white). (**C**) Tabulated results of electroporated constructs. (**D** to **I**) Confocal image montages (white hatches denote image breaks) at E6 of electroporated oculomotor nerves (green) and extraocular muscles (red) labeled with antibody to myosin [(D), (E), and (G) to (I)] or α-bungarotoxin (F); constructs as labeled. All GFP control (D), 28% of wild type (E), and only 5 to 13% of mutant α2-chimaerin electroporated oculomotor nerves extend normally from the midbrain neuroepithelium, at left, past the dorsal rectus muscle (dr), ciliary ganglion (*), and ventral (vr) and medial (mr) recti to innervate the first target, the ventral oblique (vo) muscle. Nerves expressing mutant α2-chimaerin have a higher incidence of aberrant branching [arrow in (F) with higher-magnification inset] and defasciculation than the wild type (fig. S7). Remarkably, 72% of wild-type α2-chimaerin (G), 87% of L20F α2-chimaerin (H), and 71% of G228S α2-chimaerin (I) electroporated nerves stall in the vicinity of the dr muscle. Scale bars, 200 μm; lr, lateral rectus.

β2-chimaerin (fig. S5, A and B). Both molecules are predicted to exist in inactive, closed conformations in the cytoplasm, and to unfold and translocate to the membrane in response to DAG signaling, exposing their RacGAP domains and inactivating Rac (*12*, *22*). β2-Chimaerin crystallization revealed that its inactive conformation is maintained by intramolecular interactions that impede access to the Rac and DAG binding sites (*22*). Modeling the DURS2 mutations onto the β2-chimaerin structure (fig. S5, C to E) (*22*) led to several predictions: (i) α2-Chimaerin Leu^{20} and Ile^{126} correspond to two of nine residues predicted by Canagarajah *et al*. to stabilize the β2-chimaerin closed conformation and, when mutated to alanine, were shown to enhance β2-chimaerin translocation to the membrane in vitro. (ii) Tyr^{143} is predicted to interact with Tyr^{221}, and Ala^{223} is adjacent to Asn^{224} (which is predicted to interact with Tyr^{133}); altering either of these residues may also destabilize the α2-chimaerin closed conformation. (iii) α2-Chimaerin Gly^{228} is the predicted DAG binding site. (iv) Glu^{313} is adjacent to the predicted Rac binding site. These predictions led us to hypothesize that DURS2 mutations hyperactivate α2-chimaerin RacGAP activity by destabilizing its closed conformation, or by directly altering DAG or Rac binding.

To determine whether DURS2 mutations alter the RacGAP activity of α2-chimaerin, we made full-length wild-type and mutant α2-chimaerin constructs that expressed equally stable proteins in human embryonic kidney (HEK) 293T cells and primary neurons (Fig. 3 and fig. S6A). Consistent with α2-chimaerin function, wild-type overexpression resulted in a reduction in Rac-GTP levels from baseline in HEK 293T cells (Fig. 3A). As predicted, overexpression of each mutant α2-chimaerin protein resulted in a significant further reduction in Rac-GTP levels relative to wild-type protein (Fig. 3, A and B), including when both wild-type and L20F α2-chimaerin were coexpressed together in the presence of the DAG analog, phorbol 12-myristate 13-acetate (PMA) (fig. S6, B and C). We conclude that all seven DURS2 mutations behave as dominant gain-of-function alleles (these and other data for each mutation are summarized in table S1).

Next, we quantified the amount of wild-type and mutant α2-chimaerin translocated to the HEK 293T cell membrane before and after stimulation with PMA. About 15% of wild-type α2-chimaerin, but significantly greater fractions of L20F, Y143H, A223V, and P252Q α2-chimaerin mutant proteins, translocated to the membrane fraction in a PMA dose–dependent manner (Fig. 3, C and D, and fig. S6, D and E). Thus, these four mutant residues appear to enhance membrane translocation and RacGAP activity by destabilizing the closed conformation of α2-chimaerin in response to PMA.

Individuals with DURS2-DRS harbor one mutant and one wild-type *CHN1* allele. Therefore, we performed coimmunoprecipitation experiments to investigate whether mutant hyperactivated α2-chimaerin could interact with the wild-type protein, thus potentially recruiting the wild-type pool to the membrane and further reducing Rac activity in vivo. α2-Chimaerin and each of the seven mutants were precipitated minimally by wild-type α2-chimaerin in the absence of PMA, and to a much greater extent in its presence; these results suggest that α2-chimaerin can complex with itself in a manner partially dependent on the PMA dose (Fig. 3, E and F, and fig. S6F). In addition, in the presence of PMA, the interaction of wild-type α2-chimaerin with all mutants except G228S and E313K was enhanced relative to its interaction with itself (Fig. 3F). Neither wild-type nor L20F α2-chimaerin coimmunoprecipitated with α1-chimaerin (fig. S6G); this result supports a direct or indirect association of α2-chimaerin with itself that may involve its SH2 domain.

On the basis of our findings that DURS2 mutations hyperactivate α2-chimaerin, we hypothesized that overexpressing α2-chimaerin may result in aberrant axon development in vivo. To test this idea, we used the chick in ovo system to overexpress α2-chimaerin in the embryonic oculomotor nucleus. This nucleus is more amenable than the abducens to electroporation, its development in chick has been defined (*23*), and we previously showed that some DURS2-DRS individuals have clinical and MR findings supporting a primary defect in oculomotor nerve development (*8*–*10*). Similar to rodents and humans, chick α2-chimaerin mRNA is expressed in neuroepithelia at stages of cranial motor neuron development (E4), and specifically in the developing oculomotor nucleus at the stage of axon extension and branching (E6) (Fig. 4, A and B). We electroporated embryonic chick midbrains with green fluorescent protein (GFP)–tagged wild-type and mutant α2-chimaerin (L20F and G228S) and GFP-alone control constructs at E2. These were analyzed between E5.5 (fig. S7), when oculomotor axons have extended along an unbranched trajectory to their distal target, the ventral oblique muscle (vo), and E6, when branching to the other target muscles has ensued (Fig. 4, C to I) (*23*). All 18 GFP control embryos showed a normal projection pattern in which the oculomotor nerve reached the ventral oblique muscle and branched correctly into other target muscles by E6 (Fig. 4D) (*23*). In the majority (71 to 87%) of embryos overexpressing wild-type or mutant constructs, the oculomotor nerve stalled and its axons terminated prematurely adjacent to the dorsal rectus muscle (Fig. 4, G to I). In addition, 67% of embryos overexpressing mutant constructs displayed aberrant branching and/or defasciculation of the oculomotor nerve, whereas only 24% of embryos overexpressing wild-type constructs did so (Fig. 4F and fig. S7, A to H). Regardless of the construct we used, the electroporated oculomotor nucleus appeared normal in size and neuron cell bodies displayed normal sorting, including normal migration across the midline (fig. S7, I and J) (*23*), consistent with a primary defect in axon rather than cell body development. Taken together, these observations suggest that elevated RacGAP activity as a result of hyperactivated mutant or overexpressed wild-type α2-chimaerin results in deregulation of normal oculomotor axon development.

Eph receptors and ephrins (*24*), as well as neuropilin receptors and semaphorins (*25*), are expressed in developing cranial motor nuclei in the chick and/or rodent. Several recent papers report that α2-chimaerin interacts with the EphA4 receptor and inactivates Rac in response to ephrin/EphA4 signaling (*13*–*16*). Loss of α2-chimaerin impairs EphA4 forward signaling in vivo and eliminates ephrin-induced growth cone collapse in vitro (*13*–*16*). α2-Chimaerin has also been implicated in semaphorin 3A–induced growth cone collapse (*12*). EphA4 receptor stimulation can recruit and activate phospholipase C–γ1, elevating DAG levels (*26*). Therefore, mutant α2-chimaerin may be hyperactivated in response to chemorepellents such as ephrins or semaphorins, resulting in pathological inactivation of Rac and altered transduction of downstream signals (fig. S8, A to C).

Mice with loss of α2-chimaerin have disrupted ephrin/EphA4 signaling and elevated Rac-GTP levels, with a phenotype limited to a hopping rabbit-like gait resulting from excessive and aberrant midline crossing of corticospinal tract axons and spinal interneuron projections, with no cranial nerve defects reported (*13*–*15*). We have now identified human α2-chimaerin mutations that enhance its function, reduce Rac-GTP levels, and lead to an ocular motor phenotype as a result of errors in cranial motor neuron development. It is remarkable that the up- and down-regulation of such a widely expressed signaling molecule results in two restricted and apparently non-overlapping phenotypes. It remains to be determined in which signaling pathways α2-chimaerin functions in corticospinal and cranial motor axons and why these different motor circuits are uniquely vulnerable to different perturbations in Rho GTPase activity.

References and Notes

1. See supporting material on *Science* Online.
2. E. C. Engle, *Arch. Neurol.* **64**, 633 (2007).
3. J. Jen *et al*., *Science* **304**, 1509 (2004); published online 22 April 2004 (10.1126/science.1096437).
4. M. G. Hotchkiss, N. R. Miller, A. W. Clark, W. G. Green, *Arch. Ophthalmol.* **98**, 870 (1980).
5. N. R. Miller, S. M. Kiel, W. R. Green, A. W. Clark, *Arch. Ophthalmol.* **100**, 1468 (1982).
6. B. Appukuttan *et al*., *Am. J. Hum. Genet.* **65**, 1639 (1999).
7. J. C. Evans, T. M. Frayling, S. Ellard, N. J. Gutowski, *Hum. Genet.* **106**, 636 (2000).
8. E. C. Engle, C. Andrews, K. Law, J. L. Demer, *Invest. Ophthalmol. Vis. Sci.* **48**, 189 (2007).
9. M. Chung, J. T. Stout, M. S. Borchert, *Ophthalmology* **107**, 500 (2000).
10. J. L. Demer, R. A. Clark, K. H. Lim, E. C. Engle, *Invest. Ophthalmol. Vis. Sci.* **48**, 194 (2007).
11. M. B. Dalva, *Neuron* **55**, 681 (2007).
12. M. Brown *et al*., *J. Neurosci.* **24**, 8994 (2004).
13. T. Iwasato *et al*., *Cell* **130**, 742 (2007).

14. H. Wegmeyer *et al.*, *Neuron* **55**, 756 (2007).
15. A. A. Beg, J. E. Sommer, J. H. Martin, P. Scheiffele, *Neuron* **55**, 768 (2007).
16. L. Shi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16347 (2007).
17. T. J. Van de Ven, H. M. VanDongen, A. M. VanDongen, *J. Neurosci.* **25**, 9488 (2005).
18. P. Buttery *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1924 (2006).
19. J. M. Dong, P. Smith, C. Hall, L. Lim, *Eur. J. Biochem.* **227**, 636 (1995).
20. C. Hall *et al.*, *Mol. Cell. Biol.* **13**, 4986 (1993).
21. C. Hall *et al.*, *J. Neurosci.* **21**, 5191 (2001).
22. B. Canagarajah *et al.*, *Cell* **119**, 407 (2004).
23. J. K. Chilton, S. Guthrie, *J. Comp. Neurol.* **472**, 308 (2004).
24. J. E. Cooke, C. B. Moens, *Trends Neurosci.* **25**, 260 (2002).
25. S. Guthrie, *Nat. Rev. Neurosci.* **8**, 859 (2007).
26. L. Zhou *et al.*, *J. Neurosci.* **27**, 5127 (2007).
27. M. Sahin *et al.*, *Neuron* **46**, 191 (2005).
28. We dedicate this paper to the memory of Krystal Law, who researched DURS2 in the Engle lab for her undergraduate thesis at Harvard University. We thank the families for their participation, members of the Engle lab for their thoughtful comments, J. Demer for pedigree referral, and M. Gregas, A. Di Nardo, Y. Harada, and I. Eisenberg for technical advice or assistance. Supported by grants from the National Eye Institute (E.C.E.), the Children's Hospital Boston Mental Retardation and Developmental Disabilities Research Center (E.C.E. and M.S.), the Spinal Muscular Atrophy Foundation and American Academy of Neurology (M.S.), South West Regional Development Agency (UK) (J.C. and J.A.), Wellcome Trust (M.C., N.J.G., S.G., S.L., M.P., and E.Y.), Medical Research Council (UK) (M.C., S.G., and S.L.), Clayton Foundation for Research (J.T.S. and B.A.), Research to Prevent Blindness Inc. [J.T.S., B.A., and A.I. (Career Development Award and unrestricted grant to the University of Tennessee Health Science Center, Hamilton Eye Institute)], and Futura-Onlus, Italy (A.B.). E.C.E. is a Howard Hughes Medical Institute Investigator. GenBank accession numbers: human *CHN1* mRNA, NM_001822; mouse *CHN1* mRNA, NM_001113246; chick *CHN1* mRNA, NM_001012952; human α2-chimaerin protein sequence, NP_001813. Protein Data Bank ID: human β2-chimaerin protein sequence, 1XA6.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1156121/DC1
Materials and Methods
Figs. S1 to S8
Table S1
References

5 February 2008; accepted 30 June 2008
Published online 24 July 2008;
10.1126/science.1156121
Include this information when citing this paper.

Deletional Tolerance Mediated by Extrathymic Aire-Expressing Cells

James M. Gardner,[1] Jason J. DeVoss,[1] Rachel S. Friedman,[2] David J. Wong,[3] Ying X. Tan,[1] Xuyu Zhou,[1] Kellsey P. Johannes,[1] Maureen A. Su,[1,4] Howard Y. Chang,[3] Matthew F. Krummel,[2] Mark S. Anderson[1]*

The prevention of autoimmunity requires the elimination of self-reactive T cells during their development and maturation. The expression of diverse self-antigens by stromal cells in the thymus is essential to this process and depends, in part, on the activity of the autoimmune regulator (*Aire*) gene. Here we report the identification of extrathymic *Aire*-expressing cells (eTACs) resident within the secondary lymphoid organs. These stromally derived eTACs express a diverse array of distinct self-antigens and are capable of interacting with and deleting naïve autoreactive T cells. Using two-photon microscopy, we observed stable antigen-specific interactions between eTACs and autoreactive T cells. We propose that such a secondary network of self-antigen–expressing stromal cells may help reinforce immune tolerance by preventing the maturation of autoreactive T cells that escape thymic negative selection.

Immunological tolerance to self is essential in the prevention of autoimmune disease. Mechanisms of central tolerance are mediated in part through the expression of a wide array of otherwise tissue-specific self-antigens (TSAs) such as insulin and thyroglobulin in specialized medullary thymic epithelial cells (mTECs) (*1*–*3*). The thymic expression of many of these TSAs is dependent on the autoimmune regulator (*Aire*) gene (*4*, *5*), and mutations in *Aire* lead to severe, multiorgan, tissue-specific autoimmunity in both mice (*4*, *6*) and humans (*7*, *8*). Although these results reveal a role for thymic *Aire*, self-tolerance must continue to be enforced after T cells leave the thymus. Consistent with this fact, *Aire* expression is also detectable outside the thymus, notably in the secondary lymphoid tissues (*4*, *9*), although the identity and function of such extrathymic *Aire*-expressing cells remain unclear (*10*, *11*). Here we identify a population of extrathymic *Aire*-expressing cells and examine a potential role for *Aire* in maintaining peripheral tolerance.

To accurately label *Aire*-expressing cells in vivo, we employed a bacterial artificial chromosome (BAC) transgenic approach (*12*) using the murine *Aire* locus modified to drive expression of green fluorescent protein (*Gfp*) fused to an autoimmune diabetes-related self-antigen gene, islet-specific glucose-6-phosphatase–related protein (*Igrp*) (Fig. 1A) (*13*). IGRP is a pancreatic β cell–specific protein against which autoreactive CD8 T cells are produced in both mouse and human autoimmune diabetes (*14*–*17*). We elected to include *Igrp* in our transgenic construct because it is not detectable in the thymus (fig. S1A) and because an IGRP-specific T cell receptor (TCR)–transgenic line (*8.3*) (*14*) can be used to monitor interactions of *Igrp-Gfp*–expressing cells. To verify the fidelity of the BAC transgene in recapitulating endogenous *Aire* expression in the resultant *Adig* (*Aire*-driven *Igrp-Gfp*) transgenic mice, thymic sections were co-stained for Aire and GFP, revealing thymic GFP expression highly restricted to *Aire*-expressing cells in the medulla (Fig. 1B and fig. S1B). By flow cytometry, GFP^+ cells were detectable specifically within the mTEC compartment (Fig. 1C), with approximately 30 to 40% of mTECs being GFP-positive. GFP^+ mTECs expressed uniformly high levels of class II major histocompatibility complex (MHC) and CD80 (Fig. 1D), as in previous studies of *Aire*-expressing mTECs (*3*, *11*, *18*).

To test how the introduction of IGRP into the thymic medullary epithelium might affect T cell selection, we compared *8.3* TCR-transgenic and *8.3*/*Adig* double-transgenic mice. Tetramer staining confirmed that the *8.3*/*Adig* double-transgenic mice showed a significant decrease in the percent and avidity of IGRP-specific $CD8^+$ T cells in the thymus (Fig. 1E) and in the periphery (fig. S1C). Further, although IGRP-reactive $CD8^+$ T cells were readily detected in the polyclonal wild-type NOD background, they were completely absent in *Adig* NOD mice (fig. S1D). To test the functional impact of this negative selection, *8.3* and *8.3*/*Adig* mice were followed for diabetes incidence, and *8.3*/*Adig* mice were completely protected from disease (Fig. 1F).

In a broad tissue survey using immunofluorescent anti-GFP staining, expression of the transgene was undetectable in most tissues, but distinct populations of extrathymic transgene-expressing cells were observed within the lymph nodes and spleen (fig. S2A). These extra-thymic *Aire*-expressing cells (eTACs) were generally confined to the T cell zones of the secondary lymphoid organs and preferentially localized to T cell–B cell boundary regions (Fig. 2, A and B). Immunofluorescent co-stains showed these cells to be negative for B cell (B220), fibroblastic reticular cell (gp38 and ERTR-7), and dendritic cell (CD11c) markers, but positive for class II MHC (Fig. 2, A and B). Reciprocal bone-marrow chimeras demonstrated that many of

[1]Diabetes Center, University of California San Francisco (UCSF), San Francisco, CA 94122, USA. [2]Department of Pathology, UCSF, San Francisco, CA 94143, USA. [3]Program in Epithelial Biology, Cancer Biology Program, Stanford University School of Medicine, Stanford, CA 94305, USA. [4]Department of Pediatrics, UCSF, San Francisco, CA 94122, USA.

*To whom correspondence should be addressed. E-mail: manderson@diabetes.ucsf.edu

these cells were stromal in origin (fig. S3A). Given this, we isolated secondary lymphoid stroma by flow cytometry and found that, as in the thymus, a population of GFP$^+$ cells was present that was CD45$^-$, MHC II$^+$ (Fig. 2C). These CD45$^-$ eTACs shared some characteristics with mTECs [being positive for MHC II, programmed death-1 ligand 1, and epithelial cell adhesion molecule (EpCAM)] but were distinct in that eTACs did not express the costimulatory molecules CD80 and CD86 (Fig. 2D). Although GFP$^+$ eTACs represented a significant percentage of the EpCAM$^+$ stromal cells in the periphery (8.5 ±2.4%), eTACs failed to bind the mTEC marker *Ulex europeus* agglutinin I or the fibroblastic reticular cell marker gp38, suggesting that they are distinct from previously described self-antigen–expressing stromal populations (*9*) (fig. S4). eTACs also appeared to be ubiquitous in lymphoid organs, because they were detected in mesenteric lymph nodes, Peyer's patches (fig. S2B), and the tertiary lymphoid structures that form in the infiltrated pancreatic islets of NOD mice (fig. S2C). By flow cytometry, GFP$^+$ cells were also detected in the CD45$^+$ compartment that expressed CD11c, although the level of GFP expression in these cells was significantly lower than in the CD45$^-$ cells and less enriched in *Aire* message (Fig. 2E and fig. S4).

To validate the idea that *Igrp-Gfp* transgene expression in eTACs reflected endogenous *Aire* expression, the eTAC surface markers identified in *Adig* mice (CD45$^-$, MHC II$^+$, EpCAM$^+$) were used to sort eTACs from nontransgenic mice, and *Aire* transcript was indeed found to be abundant in these cells, confirming that this stromal population expressed high levels of *Aire* (Fig. 2E). Co-staining of secondary lymphoid organs for Aire and GFP also showed Aire protein localized to perinuclear speckles within a subset of eTACs (Fig. 2F). The number of GFP$^+$ cells in which Aire protein could be detected was smaller in the periphery (24.8 ± 4.6%) than in the thymus (85.0 ± 4.5%), and the Aire staining was much weaker—near the limit of detection—which may explain why previous attempts to identify these cells in the absence of the transgenic reporter have been difficult (*11*).

Because *Aire* has been shown to play an important role in the transcriptional regulation of self-antigens in mTECs, we sought to define its function as a transcriptional regulator in eTACs. GFP$^+$ eTACs were sorted for microarray analysis from the spleens and pooled lymph nodes of *Adig* mice crossed onto the *Aire*$^{+/+}$ or *Aire*$^{-/-}$ background. As in mTECs, the number of genes up-regulated by *Aire* in eTACs was greater than the number down-regulated (Fig. 3A and tables S1 and S2). Both the total number of *Aire*-regulated genes in eTACs and the fold change of expression of those genes were smaller than has been observed in mTECs, perhaps reflecting the lower and potentially transient expression of *Aire* in the periphery. There was little overlap between *Aire*-regulated genes in eTACs and those in mTECs (Fig. 3B), suggesting that *Aire* in the periphery may regulate the expression of a distinct set of self-antigens. Despite these differences, however, we found a significant enhancement for TSAs among the positively *Aire*-regulated genes in eTACs (Fig. 3C), several of which were confirmed by quantitative reverse transcription polymerase chain

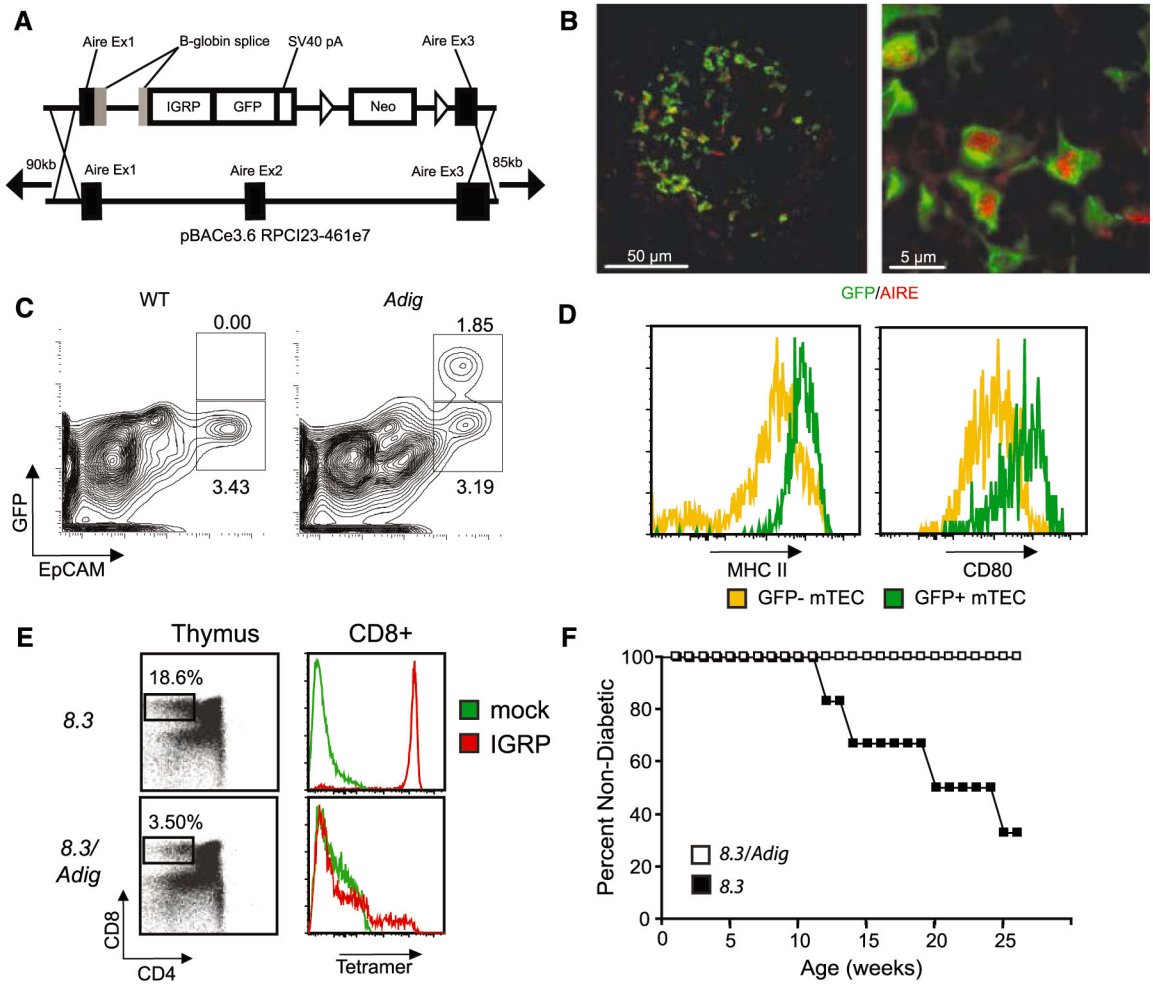


Fig. 1. The *Adig* transgene recapitulates *Aire* expression in the thymus and mediates the negative selection of autoreactive T cells. (**A**) Schematic of *Igrp-Gfp* transgene targeting the *Aire* BAC. Targeting replaces *Aire* exon 2, the coding portion of exon 1, and part of exon 3, so that the transgene does not make functional Aire. (**B**) Immunofluorescent staining of GFP (green) and Aire (red) in thymic frozen sections of *Adig* mice. (**C**) Flow cytometry of CD45-, 4′,6′-diamidino-2-phenylindole– (DAPI-), and Ly51lo-gated thymic stromal cells from wild-type (left) and *Adig* (right) NOD mice. (**D**) Flow cytometry of CD45-, DAPI-, and Ly51lo-gated EpCAM$^+$ mTECs from *Adig* NOD mice gated as GFP$^-$ (yellow) or GFP$^+$ (green) and stained for MHC II (left) and CD80 (right). (**E**) Thymocytes from *8.3* TCR-transgenic mice (top panels) or double-transgenic *8.3/Adig* mice (bottom panels) stained for CD4/CD8 (left panels), or pre-gated as CD4$^-$ CD8$^+$ and stained for IGRP-mimetope (red) or mock (green) peptide/K^d tetramer reactivity (right panels). (**F**) Diabetes incidence curves for *8.3* TCR-transgenic mice (n = 6) and *8.3/Adig* double-transgenic mice (n = 10).

reaction (PCR) (Fig. 3D). The list of genes regulated by *Aire* in eTACs also included a number of self-antigens whose human homologs have been described as autoantigens in human autoimmune diseases, including desmoglein 1a (pemphigus foliaceus) (*19*), ladinin 1 (linear IgA dermatosis) (*20*), and the *N*-methyl-D-aspartate (NMDA) receptor 2C (systemic lupus erythematosis) (*21*). Like other professional antigen-presenting cells (APCs), eTACs also expressed a large number of antigen-processing and presentation genes, suggesting a likely role for T cell interaction (Fig. 3E). Comparison of global gene expression profiles between eTACs, mTECs, cortical thymic epithelial cells (cTECs), thymic dendritic cells (DCs), and macrophages indicated that eTACs were most similar to DCs and mTECs (Fig. 3F).

To directly test the ability of eTACs to promote tolerance by interacting with and deleting autoreactive T cells, adoptive co-transfer of carboxyfluorescein diacetate succinimidyl ester (CFSE)–labeled congenic *8.3* and polyclonal $CD8^+$ T cells was used. Upon transfer into wild-type hosts, *8.3* $CD8^+$ T cells proliferated only in the pancreatic lymph nodes and persisted in all lymph nodes for up to 2 weeks (Fig. 4A). When transferred into *Adig* hosts, however, the entire population of *8.3* T cells had proliferated rapidly in all secondary lymphoid organs by 3 days after transfer (Fig. 4A) and had nearly disappeared by 2 weeks after transfer (Fig. 4, A and C). To confirm that the absence of *8.3* T cells in transgenic recipients was due to cell death and not egress, these experiments were repeated in the presence of the S1P1 inhibitor FTY 720 (*22*) (fig. S5). Further, to determine whether a stromal eTAC population was sufficient to directly mediate this deletion, irradiated wild-type and *Adig* mice were reconstituted with bone marrow deficient in β*2-microglobulin* ($\beta 2M^{-/-}$), so that only stromal cells were capable of interacting with $CD8^+$ lymphocytes. Chimerism was confirmed both by blood typing and functionally by the observation that *8.3* T cells failed to proliferate in the pancreatic lymph nodes of nontransgenic $\beta 2M^{-/-}$ reconstituted mice (Fig. 4B). In contrast, *8.3* T cells continued to proliferate in all secondary lymphoid organs of *Adig* $\beta 2M^{-/-}$ reconstituted mice at 3 days, and all divided *8.3* T cells had been deleted by 14 days (Fig. 4, B and C). Although *8.3* cell division was less robust at day 3 in $\beta 2M^{-/-}$ chimeric *Adig* mice when compared to unirradiated *Adig* mice, antigen-specific cell death at day 14 was even more dramatic in this setting (Fig. 4, A versus B).

To clearly delineate whether eTACs can directly interact with T cells, two-photon microscopy of explanted lymph nodes was used. GFP^+ cells were observed in all examined lymph nodes of *Adig* mice (movie S1), in a distribution that mirrored the localization of eTACs observed by immunofluorescent staining (movie S2 and Fig. 2A). Adoptive cotransfer of fluorophore-labeled *8.3* and polyclonal $CD8^+$ T cells demonstrated sustained antigen-specific association between naïve *8.3* T cells and eTACs as early as 4 hours after transfer (Fig. 4, D and E, and movies S3 and S4). *8.3* T cells exhib-

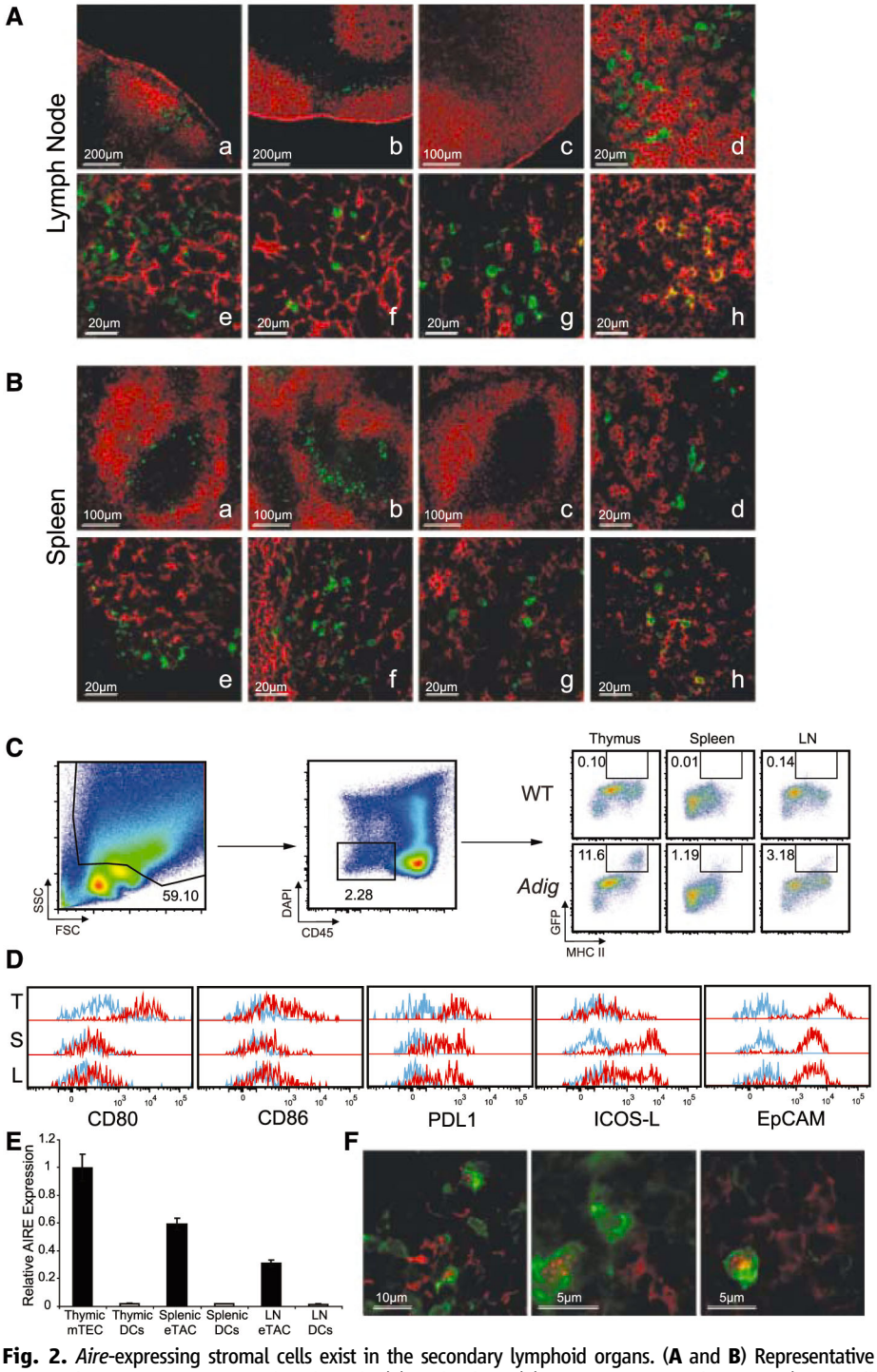


Fig. 2. *Aire*-expressing stromal cells exist in the secondary lymphoid organs. (**A** and **B**) Representative immunofluorescent co-stains of lymph node (A) and spleen (B) sections co-stained for GFP (green, all sections) and B220 (a to d), gp38 (e), ERTR-7 (f), CD11c (g), or MHC II (h; all red). Images a, b, and d to h are from *Adig* NOD mice; images labeled c are from wild-type (WT) NOD mice. (**C**) Flow cytometric analysis and gating from *Adig* and wild-type NOD thymus, spleen, and lymph node stroma analyzed for CD45, DAPI, MHC II, and GFP. (**D**) Flow cytometric analysis of *Adig* NOD thymus (T), spleen (S), and lymph nodes (L), showing expression of indicated markers (red) or isotype staining (blue) in mTECs and eTACs respectively, defined as $CD45^-$, $DAPI^-$, MHC II^+, GFP^+. (**E**) Real-time PCR analysis of *Aire* expression relative to endogenous control in cell-sorted mTECs and eTACs [$CD45^-$, propidium iodide–negative (PI^-), MHC II^+, $CD11c^-$, $EpCAM^+$] and DCs ($CD45^+$, PI^-, MHC II^+, $CD11c^+$, $EpCAM^-$) of nontransgenic NOD thymus, spleen, and lymph node. (**F**) Immunofluorescent GFP (green) and Aire (red) co-stains of lymph nodes from *Adig* NOD mice.

ited distinct reductions in speed and total displacement relative to polyclonal CD8$^+$ T cells (Fig. 4, D and F), and *8.3* T cells spent significantly more time both stopped and in direct contact with eTACs (Fig. 4F). Indeed, the upper limit of this interaction time is unknown, as many *8.3* T cells spent the entire 30-min duration of the acquisition periods in continuous contact (Fig. 4G and movie S5). Some GFP$^+$ cells in the lymph node appeared highly motile, but *8.3* T cells were able to maintain antigen-specific interactions despite this motility (Fig. 4E and movie S6). Together, these results suggested that eTACs can form early, stable, long-term contacts with naïve autoreactive T cells entering the lymph node, and that such interaction leads to rapid proliferation and deletion of these T cells.

We have identified a previously unknown population of extrathymic *Aire*-expressing cells (eTACs) that may play an important and previously uncharacterized role in self-tolerance via the deletion of autoreactive T cells. eTACs share certain characteristics with mTECs, including being equipped to act as professional APCs and the Aire-regulated expression of TSAs. The set of Aire-regulated TSAs expressed in eTACs appears to have little overlap with thymic Aire-regulated antigens, which may explain why previous efforts examining known thymic TSAs for expression in secondary lymphoid organs have been inconsistent or conflicting. The lack of overlap suggests that there may be a higher order of Aire-dependent transcriptional regulation of TSA expression, whether direct or indirect, that differs between the thymus and the periphery. Many important questions remain, including the developmental origin of eTACs and their relationship with other possible stromal APC populations (*9*). In this regard, the precise identification of eTACs will provide the framework for exploring these issues. Further, it will be important to determine the physiologic relevance of this cell population in a nontransgenic setting, given the

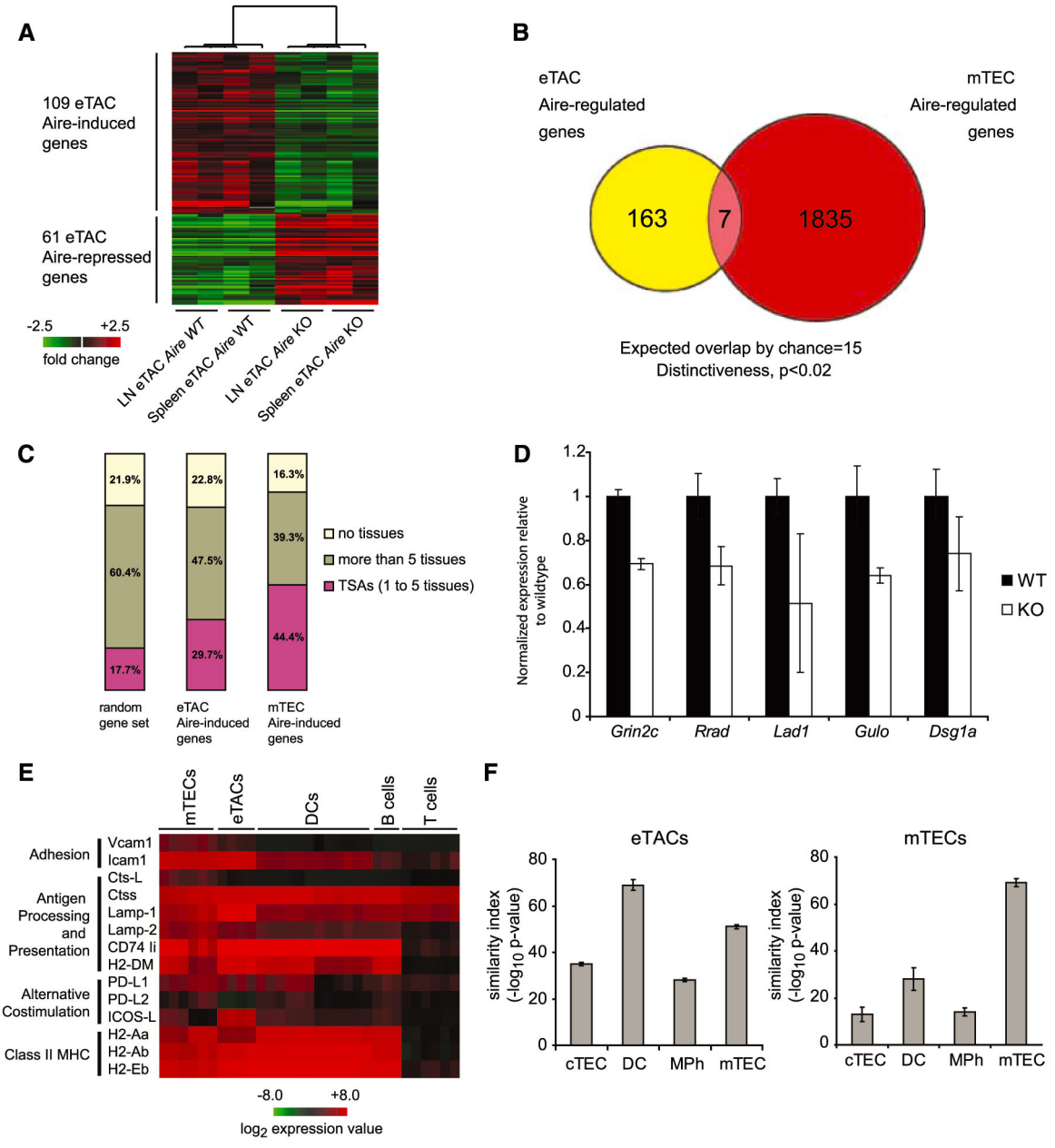


Fig. 3. *Aire* regulates the expression of a set of tissue-specific antigens in eTACs. (**A**) Heat map and unsupervised clustering of *Aire*-regulated genes in eTACs. Pooled eTACs were sorted from lymph nodes and spleens from cohorts of 3- to 6-week-old *Adig Aire*$^{+/+}$ and *Adig Aire*$^{-/-}$ NOD mice. Each of the eight arrays represents three to five pooled mice. (**B**) Schematic diagram of the unshared and common genes regulated by *Aire* in eTACs and mTECs. (**C**) Classification of *Aire*-regulated genes in eTACs based on tissue specificity, as compared to mTECs and to a random gene set. (**D**) Real-time PCR analysis of *Aire*-regulated TSAs in eTACs, normalized to endogenous control. eTACs were sorted from pooled nontransgenic *Aire*$^{+/+}$ [black bars, wild type (WT)] and *Aire*$^{-/-}$ [white bars, knockout (KO)] NOD spleens based on the surface marker profile CD45$^-$, PI$^-$, CD11c$^-$, MHC II$^+$, EpCAM$^+$, and characterized for expression of glutamate receptor NMDA2C (*Grin2c*), Ras-related associated with diabetes (*Rrad*), ladinin 1 (*Lad1*), gulonolactone (L-) oxidase (*Gulo*), and desmoglein 1 alpha (*Dsg1a*). (**E**) Expression of antigen-processing and -presentation genes in eTACs relative to other lymphoid cell populations after median-centered normalization to the expression of all genes in each array. (**F**) Global expression profile similarity of eTACs (left) and mTECs (right) to other relevant cell types based on Pearson correlation values calculated for population-specific centroids.

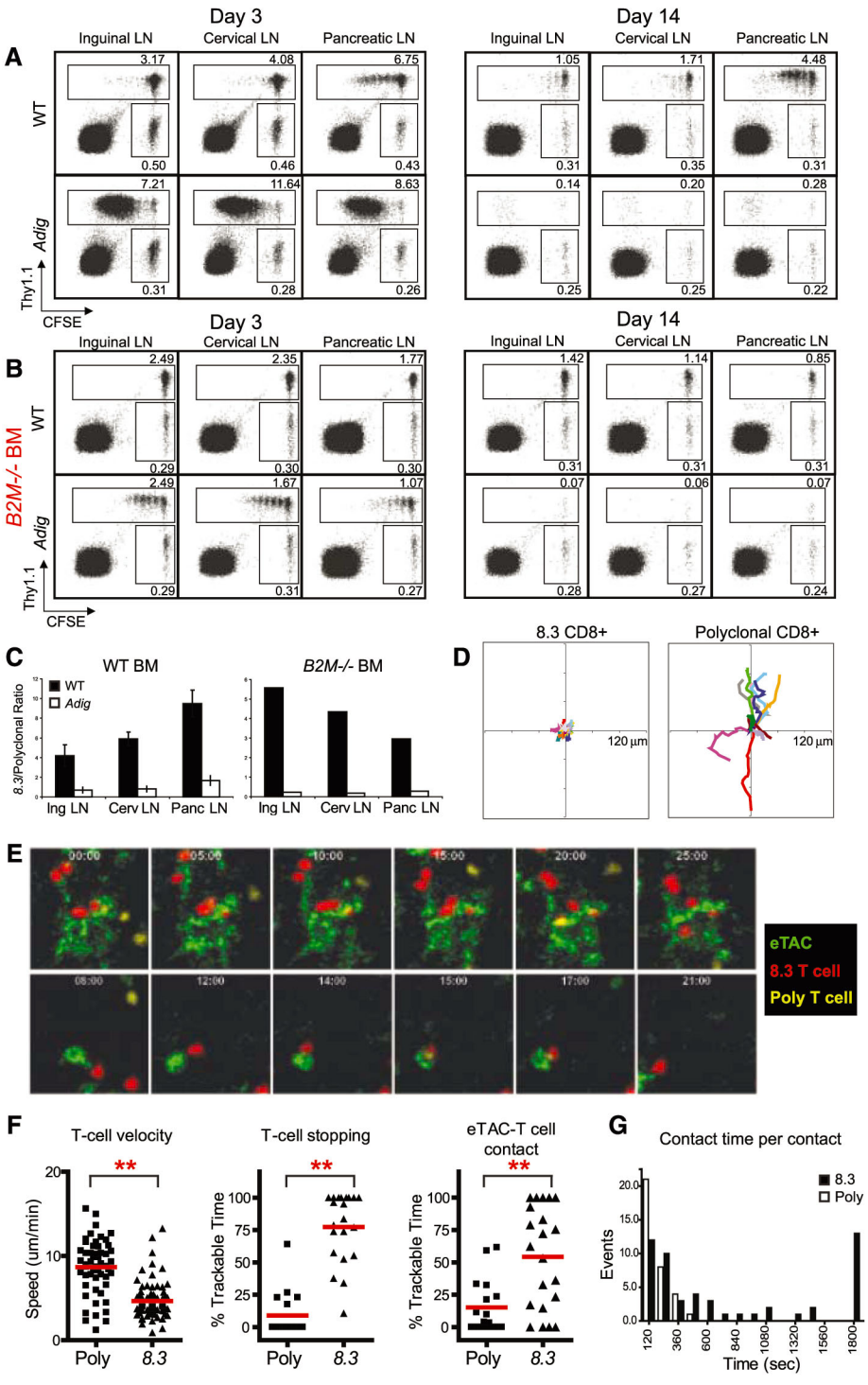


Fig. 4. eTACs directly interact with autoreactive lymphocytes and mediate deletional tolerance. (**A**) Flow cytometry of CFSE-labeled and adoptively cotransferred *8.3* $CD8^+$ T cells (Thy1.1) and polyclonal $CD8^+$ T cells (Thy 1.2). Cells were harvested at day 3 (left) and day 14 (right) after transfer (**B**) Adoptive transfer of the same donor populations as (A) into lethally irradiated wild-type (top) and *Adig* (bottom) recipients reconstituted with $\beta 2M^{-/-}$ bone marrow. (**C**) Quantitation of antigen-specific deletion after adoptive transfer at day 14 in (A) and (B), showing the ratio of *8.3* to polyclonal $CD8^+$ T cells in wild-type (black) and *Adig* (white) NOD recipients. Bars are representative of at least three mice each. (**D** to **G**) Two-photon imaging of *8.3* and polyclonal $CD8^+$ T cells in axillary lymph nodes 4 hours after adoptive transfer into *Adig* NOD recipients. (D) 10-min displacement analysis of all *8.3* $CD8^+$ T cell tracks (left) and polyclonal $CD8^+$ T cell tracks (right). (E) Representative images of interaction between *8.3* $CD8^+$ T cells (red), polyclonal $CD8^+$ T cells (yellow) and eTACs (green). (F) Average T cell track speed (left), percent of time in which a T cell is stopped (middle), and percent of time in which a T cell is making contact with an eTAC (right) among polyclonal (Poly) and *8.3* (*8.3*) CD8 T cells. $^{**}P < 0.001$. (G) Histogram displaying the duration of individual T cell–eTAC interaction times per contact for polyclonal (black bars) and *8.3* (white bars) $CD8^+$ T cells.

modest levels of TSA expression and *Aire*-dependent gene up-regulation in eTACs (approximately twofold). However, it is notable that similar analyses of *Aire*-expressing mTECs have demonstrated that TSAs are expressed at low levels in these cells (*2*), but even this low level of expression has proven critical for the maintenance of immune tolerance (*4*, *23*). Our findings suggest that eTACs may represent a safety net within the entire immunologic periphery, which functions to screen out naïve autoreactive T cell clones that escape thymic negative selection. Finally, we speculate that eTACs may play an increasingly important role with advancing age, as the thymus involutes and the burden of maintaining self-tolerance shifts to the periphery.

References and Notes

1. K. M. Smith, D. C. Olson, R. Hirose, D. Hanahan, *Int. Immunol.* **9**, 1355 (1997).
2. J. Derbinski, A. Schulte, B. Kyewski, L. Klein, *Nat. Immunol.* **2**, 1032 (2001).
3. J. Derbinski *et al.*, *J. Exp. Med.* **202**, 33 (2005).
4. M. S. Anderson *et al.*, *Science* **298**, 1395 (2002).
5. A. Liston, S. Lesage, J. Wilson, L. Peltonen, C. C. Goodnow, *Nat. Immunol.* **4**, 350 (2003).
6. C. Ramsey *et al.*, *Hum. Mol. Genet.* **11**, 397 (2002).
7. K. Nagamine *et al.*, *Nat. Genet.* **17**, 393 (1997).
8. Finnish-German APECED Consortium, *Nat. Genet.* **17**, 399 (1997).
9. J. W. Lee *et al.*, *Nat. Immunol.* **8**, 181 (2007).
10. M. Halonen *et al.*, *J. Histochem. Cytochem.* **49**, 197 (2001).
11. F. X. Hubert *et al.*, *J. Immunol.* **180**, 3824 (2008).
12. W. Yu *et al.*, *Science* **285**, 1080 (1999).
13. Materials and methods are available as supporting material on *Science* Online.
14. M. Nagata, P. Santamaria, T. Kawamura, T. Utsugi, J. W. Yoon, *J. Immunol.* **152**, 2042 (1994).
15. B. Han *et al.*, *J. Clin. Invest.* **115**, 1879 (2005).
16. S. M. Lieberman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8384 (2003).
17. I. Jarchum, L. Nichol, M. Trucco, P. Santamaria, T. P. Dilorenzo, *Clin. Immunol.* **127**, 359 (2008).
18. D. Gray, J. Abramson, C. Benoist, D. Mathis, *J. Exp. Med.* **204**, 2521 (2007).
19. M. S. Lin *et al.*, *J. Clin. Invest.* **105**, 207 (2000).
20. M. P. Marinkovich, T. B. Taylor, D. R. Keene, R. E. Burgeson, J. J. Zone, *J. Invest. Dermatol.* **106**, 734 (1996).
21. C. Kowal *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19854 (2006).
22. M. Matloubian *et al.*, *Nature* **427**, 355 (2004).
23. J. DeVoss *et al.*, *J. Exp. Med.* **203**, 2727 (2006).

24. We thank J. Bluestone, A. Abbas, N. Killeen, and J. Cyster for helpful discussion. J.G. is funded by the UCSF/NIH Medical Scientist Training Program and the American Diabetes Association, J.D. by the Giannini Foundation, R.F. by the Larry L. Hillblom Foundation, and D.W. by a Dermatology Foundation Research Career Development Award. M.S.A. is supported by the Pew Scholars Program, the Sandler Foundation, and the Burroughs Wellcome Fund. This work was supported in part by NIH (to M.S.A.). The authors have no competing financial interests.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5890/843/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 and S2
References
Movies S1 to S6

21 April 2008; accepted 14 July 2008
10.1126/science.1159407

Dichotomous Dopaminergic Control of Striatal Synaptic Plasticity

Weixing Shen,[1] Marc Flajolet,[2] Paul Greengard,[2] D. James Surmeier[1]*

At synapses between cortical pyramidal neurons and principal striatal medium spiny neurons (MSNs), postsynaptic D1 and D2 dopamine (DA) receptors are postulated to be necessary for the induction of long-term potentiation and depression, respectively—forms of plasticity thought to underlie associative learning. Because these receptors are restricted to two distinct MSN populations, this postulate demands that synaptic plasticity be unidirectional in each cell type. Using brain slices from DA receptor transgenic mice, we show that this is not the case. Rather, DA plays complementary roles in these two types of MSN to ensure that synaptic plasticity is bidirectional and Hebbian. In models of Parkinson's disease, this system is thrown out of balance, leading to unidirectional changes in plasticity that could underlie network pathology and symptoms.

The striatal release of dopamine (DA) is intimately linked to associative learning and habit formation (*1*, *2*). This role is thought to be mediated by controlling corticostriatal synaptic plasticity (*3*–*5*). However, efforts to characterize how this control is exerted have met with only modest success. Principal striatal medium spiny neurons (MSNs) are heterogeneous in their expression of DA receptors (*6*), falling into one of two equally sized, morphologically similar groups. One group expresses predominantly D1 DA receptors (D1 MSNs), whereas the other expresses D2 DA receptors (D2 MSNs). These two receptors appear to modulate long-term changes in glutamatergic synaptic plasticity in MSNs in different ways. D1 DA receptor signaling promotes long-term potentiation (LTP) (*3*, *7*), whereas D2 DA receptor signaling promotes long-term depression (LTD) (*8*, *9*). In animal models of Parkinson's disease (PD), where striatal DA levels are very low, both forms of synaptic plasticity in MSNs appear to be lost, which suggests that DA receptor signaling is necessary for their induction (*7*, *8*).

This simple picture poses an obvious conceptual puzzle. If DA is necessary for the induction of synaptic plasticity and D1 and D2 receptors are expressed by different MSNs, then synaptic plasticity must be unidirectional in each population. However, a high percentage of MSNs display both forms of plasticity (*3*, *7*, *9*).

In an attempt to resolve this paradox, we reexamined glutamatergic synaptic plasticity in brain slices from transgenic mice in which the expression of D1 or D2 receptors was reported by coexpression of green fluorescent protein (GFP). GFP expression in these mice faithfully reports MSN phenotype, allowing D1 and D2 receptor–expressing MSNs to be reliably sampled (*10*, *11*). We induced plasticity by pairing afferent stimulation with postsynaptic spikes in short bursts that were repeated at a theta frequency (5 Hz). At most synapses, a Hebbian form of spike-timing-dependent plasticity (STDP) is induced by this protocol (*12*). That is, when presynaptic activity precedes postsynaptic spiking, LTP is induced, whereas reversing the order induces LTD (*12*–*14*). To induce STDP, glutamatergic afferent fibers were stimulated with a small pipette close (~100 μm) to the soma of an identified MSN that was driven to spike by current injected through a somatic perforated membrane patch (Fig. 1, A to C) (*12*, *13*, *15*).

In D2 MSNs (fig. S1), repeated pairing of a synaptic stimulation with a postsynaptic spike 5 ms later resulted in LTP of the synaptic response (Fig. 1D). In contrast, preceding synaptic stimulation (–10 ms) with a short burst of postsynaptic spikes induced LTD (Fig. 1E). There were no lasting alterations in synaptic strength with unpaired presynaptic or postsynaptic activity (fig. S1).

Previous studies of striatal LTD induced by conventional plasticity protocols have underscored the importance of D2 receptors (*7*, *8*, *16*). In D2 MSNs, timing-dependent LTD was disrupted by antagonizing D2 receptors with sulpiride (control $n = 5$; sulpiride $n = 5$; $P < 0.05$, Mann-Whitney rank sum test), suggesting a similar involvement of D2 receptors (Fig. 1F). Moreover, LTD was disrupted by antagonizing CB1 endocannabinoid (fig. S2) or mGluR5 glutamate receptors (fig. S3). The combination of presynaptic activity and activation of terminal CB1 receptors leads to a lasting reduction in glutamate release probability underlying conventional LTD (*17*). Indicating a presynaptic expression mechanism, LTD was accompanied by increased trial-to-trial variation in excitatory postsynaptic potential (EPSP) amplitude (fig. S4).

The determinants of LTP at glutamatergic synapses of MSNs are less well characterized. D1 receptors are thought to be important, but they are not expressed by D2 MSNs. Moreover, blocking D1 receptors with SCH23390 had no effect on timing-dependent LTP in D2 MSNs (fig. S5). Adenosine A2a receptors, which couple to the same second messenger cascades as D1 receptors, are robustly and selectively expressed by D2 MSNs (*18*). Antagonizing these receptors—not D1 receptors—disrupted the induction of LTP in D2 MSNs (control $n = 11$; SCH58261 $n = 7$; $P < 0.001$, Mann-Whitney test) (Fig. 1G). Blocking *N*-methyl-D-aspartate (NMDA) receptors also prevented the induction of timing-dependent LTP, as with conventional LTP (*7*) [D,L-2-amino-5-phosphonovaleric acid (APV) and MK-801 $n = 5$; $P < 0.01$, Mann-Whitney test] (Fig. 1G). Furthermore, the variation in EPSP amplitude was unchanged after LTP induction, which suggests that the expression of plasticity was postsynaptic (fig. S4).

Bidirectional STDP depends upon opponent processes controlling LTD and LTP (*13*, *19*, *20*). The net change in synaptic strength is hypothesized to reflect the interaction between cellular processes sensitive to the timing of pre- and postsynaptic activity (e.g., NMDA receptors and L-type Ca^{2+} channels) and G-protein coupled receptor (GPCR)–regulated intracellular signaling cascades. By altering the balance between GPCR cascades controlling plasticity, the timing dependence of STDP can be diminished (*19*, *20*). We therefore elevated D2 receptor stimulation by bath application of quinpirole during pairing of presynaptic stimulation with a trailing (+5 ms) postsynaptic spike, a protocol that normally induces LTP. This resulted in a robust LTD ($n = 6$; $P < 0.05$, Wilcoxon signed rank test) (Fig. 1, H and I). Boosting A2a receptor signaling by bath application of CGS21680 restored LTP ($n = 6$; $P < 0.05$, Wilcoxon test) (Fig. 1, H and I). Conversely, bath application of the A2a receptor agonist in the presence of a D2 receptor antagonist led to the induction of LTP, even when postsynaptic spiking preceded presynaptic stimulation (Fig. 1F).

When the same STDP protocols were applied to D1 MSNs (fig. S1), a different picture emerged. Pairing presynaptic activity with a trailing postsynaptic spike (+5 ms) induced a robust LTP (Fig. 2A). LTP was dependent upon NMDA receptors (control $n = 5$; APV and MK-801 $n = 4$; $P < 0.05$, Mann-Whitney test) (Fig. 2B) and appeared to be postsynaptically expressed (fig. S4). However, when presynaptic activity followed postsynaptic spiking (–10 ms), EPSP amplitude did not change (Fig. 2C). The absence of LTD in D1 MSNs was consistent with recent work using a conventional plasticity protocol and has been attributed to the failure to generate endocannabinoids postsynaptically during induction, rather than to the absence of presynaptic CB1 receptors (*8*), an inference consistent with the inability of AM-251 to affect the response (control $n = 6$; AM-251 $n = 5$; $P > 0.05$, Mann-Whitney test) (Fig. 2D). We reasoned that this failure could be due to the activation of the GPCR signaling responsible for LTP induction. To test this hypothesis, D1 receptors—commonly believed to be necessary for LTP induction in MSNs (*3*, *7*)—were blocked by bath application of SCH23390 and the protocol repeated. In the absence of D1 receptor activity, pairing postsynaptic spiking with a trailing presynaptic volley led to a robust LTD ($n = 6$; $P < 0.05$, Wilcoxon test) (Fig. 2D) that was blocked by an mGluR5 antagonist (fig. S3). However, antagonizing D2 receptors did not disrupt LTD induction in D1

[1]Department of Physiology, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA. [2]Laboratory of Molecular and Cellular Neuroscience, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA.

*To whom correspondence should be addressed. E-mail: j-surmeier@northwestern.edu

MSNs (fig. S6), which suggests that D2 receptor sensitive interneuronal signaling was not engaged by local, minimal stimulation (*16*).

To determine whether attenuating D1 receptor signaling altered the timing dependence of plasticity, the effect of the positive timing protocol that normally induced a robust LTP (Fig. 2E) was reexamined in the presence of a D1 receptor antagonist. This not only prevented LTP induction, it led to the induction of LTD (SCH23390 $n = 6$; $P < 0.05$, Wilcoxon test) that was dependent upon endocannabinoid CB1 receptors (SCH23390 $n = 6$; SCH23390 and AM-251 $n = 5$; $P < 0.01$, Mann-Whitney test) (Fig. 2E, F), establishing a mechanistic parallel to LTD in D2 MSNs. Antagoniz-

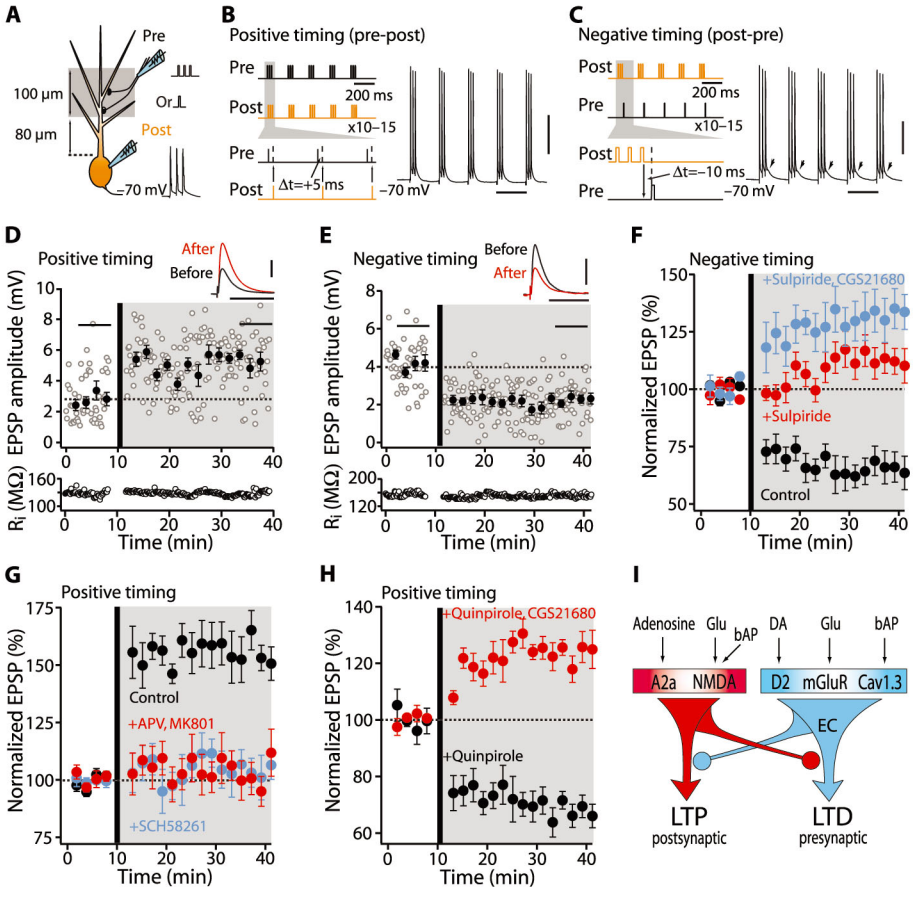


Fig. 1. D2 MSNs displayed bidirectional Hebbian STDP dependent upon D2 and A2a receptors. (**A**) Schematic illustration of the recording/stimulation configuration. (**B** and **C**) The theta-burst pairing protocols for induction of (B) LTP and (C) LTD. Scale bars, 40 mV × 200 ms. (**D**) LTP induced by a positive timing pairing. Plots show EPSP amplitude and input resistance as a function of time. The dashed line shows the average EPSP amplitude before induction. The induction was performed at the vertical bar. Filled symbols show the averages of 12 trials (±SEM). The averaged EPSP traces before and after induction are shown at the top. Scale bars, 2 mV × 100 ms. (**E**) LTD induced by a negative timing pairing. Plots and EPSP traces as in (D). Scale bars, 2 mV × 100 ms. (**F**) In the presence of D2 receptor antagonist sulpiride (10 μM), negative timing pairing failed to alter EPSP amplitude, but coapplication of A2a adenosine receptor agonist CGS 21680 (100 nM) and sulpiride led to LTP ($n = 6$; $P < 0.05$, Wilcoxon). (**G**) LTP induction ($n = 11$; $P < 0.01$, Wilcoxon) was disrupted by the NMDA receptor antagonists APV (50 μM) and MK-801 (20 μM) or the A2a receptor antagonist SCH58261 (100 nM). (**H**) In the presence of D2 receptor agonist quinpirole (10 μM, $n = 6$), the application of the positive timing protocol leads to induction of LTD. Application of quinpirole and CGS21680 (100 nM, $n = 6$) together restored LTP with a positive timing pairing. (**I**) Schematic illustration shows that activation of A2a and NMDA receptors leads to LTP and that activation of D2 and mGluR5 receptors and Cav1.3 channels leads to LTD. Moreover, A2a and D2 receptor activation oppose each other in inducing plasticity. Glu, glutamate; EC, endocannabinoid.

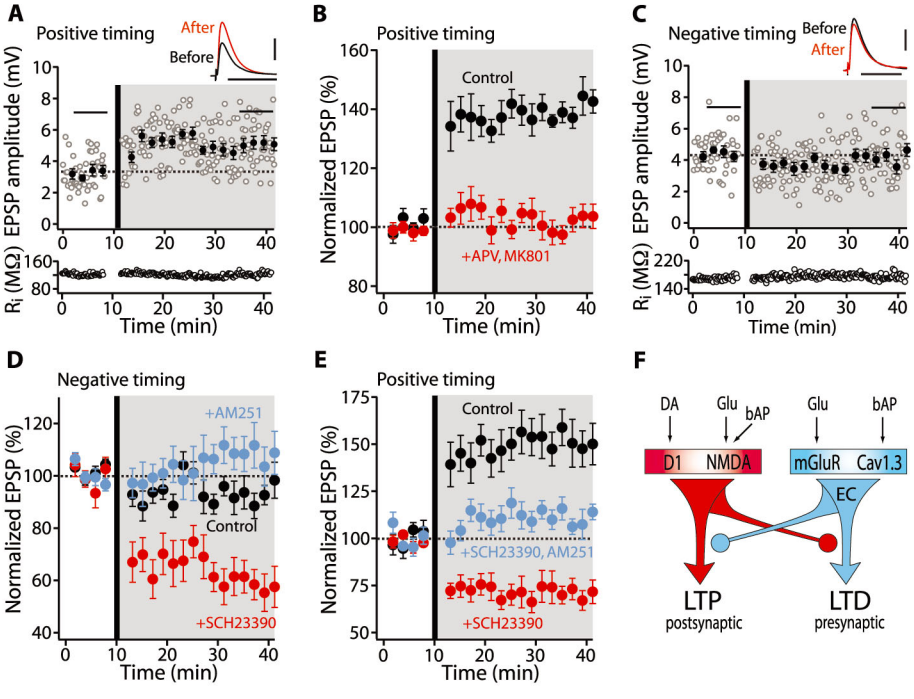


Fig. 2. D1 MSNs displayed bidirectional Hebbian STDP dependent upon D1 receptors. (**A**) LTP induction by a positive timing pairing protocol. EPSP amplitude and input resistance of the recorded cell were plotted as a function of time. The averaged EPSP traces before and after induction are shown at the top. Scale bars, 2 mV × 100 ms. (**B**) LTP induction ($n = 10$; $P < 0.01$, Wilcoxon test) was blocked by APV (50 μM) and MK-801 (20 μM). (**C**) LTD was not induced in D1 neurons with a negative pairing. Plots and EPSP traces are from a single cell, as in (A). Scale bars, 2 mV × 100 ms. (**D**) In the presence of D1 receptor antagonist SCH23390 (3 μM), a negative timing pairing revealed LTD, but in the presence of CB1 receptor antagonist AM-251 (2 μM), negative pairing failed to alter EPSP amplitude. (**E**) LTP induced by a positive timing pairing was blocked by SCH23390, revealing LTD. LTD induced in the presence of SCH23390 was disrupted by AM-251. (**F**) Schematic drawing shows that activation of D1 and NMDA receptors evokes LTP and that activation of mGluR5 receptor and Cav1.3 channels evokes LTD. Moreover, D1 and mGluR5 receptor activation oppose each other in inducing plasticity.

ing downstream presynaptic CB1 receptors alone did not alter LTP induction (fig. S7).

These experiments show that while DA makes STDP in striatal MSNs bidirectional and Hebbian (*12*), it is not necessary for the induction of synaptic plasticity. This stands in contrast to previous work asserting that DA is essential for plasticity and that striatal DA depletion in animal models of PD eliminates both LTD and LTP at MSN glutamatergic synapses (*7*, *8*). To revisit this issue, two well-established mouse models of PD were examined: In one, DA neurons were lesioned by injecting the toxin 6-hydroxydopamine (6-OHDA) into the medial forebrain bundle (Fig. 3A), and in the other, DA was depleted by disrupting the monoamine vesicular transporter with reserpine (*15*). In both models, striatal DA levels were dramatically reduced and STDP altered in the same way. In D2 MSNs, LTP was induced not only by the usual pairing protocol (6-OHDA $n = 6$; $P < 0.05$, Wilcoxon test) (Fig. 3B) (reserpine $n = 6$; $P < 0.05$, Wilcoxon test) (Fig. 3C) but also when spiking preceded synaptic stimulation, a protocol that normally induced LTD ($n = 10$; $P < 0.05$, Wilcoxon test) (Fig. 3D). As in normal tissue, this LTP was sensitive to A2a receptor block (reserpine $n = 6$; SCH58261 $n = 4$; $P < 0.05$, Mann-Whitney test) (Fig. 3C). Reestablishing D2 receptor activity with exogenous quinpirole rescued LTD ($n = 6$; $P < 0.05$, Wilcoxon test) (Fig. 3D). In contrast, LTP was not induced in D1 MSNs by pairings in which synaptic activity preceded postsynaptic spiking; rather, this protocol induced a robust LTD (6-OHDA $n = 6$; $P < 0.05$, Wilcoxon test) (Fig. 3E) (reserpine $n = 6$; $P < 0.05$, Wilcoxon test) (Fig. 3F) that was sensitive to CB1 receptor block (reserpine $n = 6$; AM-251 $n = 4$; $P < 0.01$, Mann-Whitney test) (Fig. 3F). Reestablishing D1 receptor activity with exogenous SKF81297 rescued LTP with this protocol ($n = 6$; $P < 0.05$, Wilcoxon test) (Fig. 3F).

Our studies demonstrate that DA is critical for the induction of bidirectional, timing-dependent (Hebbian) plasticity at glutamatergic synapses formed on striatal MSNs. This finding is consistent with a recent study (*14*) but conflicts with another (*21*). The discrepancy between studies could be attributable to heterogeneity in glutamatergic synapses (*22*) or the engagement of striatal interneurons capable of modulating the induction of plasticity (*7*, *16*). Although the role of these factors is important to sort out, our goal in using focal stimulation near synaptic sites was to isolate the direct actions of DA on the induction of plasticity at glutamatergic synapses of largely cortical origin. This strategy revealed that the type of DA receptor present at the postsynaptic site governed the actions of DA. Furthermore, the signaling determinants of STDP resolved in this way were very similar to those reported for conventional plasticity in MSNs (*7*–*9*), which suggests that they engaged the same cellular machinery. This insight and the recognition that DA does not act alone in regulating the induction of plasticity, but in concert with glutamate and adenosine, resolves the conflict posed by the apparent obligatory roles of D1 and D2 receptors in the induction of plasticity and their segregation in different MSN classes. It also shows how the activity of DA neurons might serve to reshape the striatal network during associative learning (*4*). In the absence of behaviorally important stimuli, DA neurons spike autonomously to maintain striatal DA concentrations at levels sufficient to keep high-affinity D2 DA receptors active, but not low affinity D1 DA receptors (*23*, *24*)—in principle enabling bidirectional, Hebbian plasticity in D2 MSNs, but not in D1 MSNs, where the low level of D1 receptor activity should permit only LTD. However, when behaviorally important stimuli drive phasic spiking of mesencephalic DA neurons, striatal DA levels rise transiently and activate D1 DA receptors (*23*); this should enable the induction of Hebbian LTP in D1 MSNs. This stark dichotomy provides a cellular foundation for the view that the networks anchored by these two MSNs regulate distinct aspects of behavior and learning (*25*, *26*).

The existence of opponent processes regulating the induction of plasticity also has implications for disease states where DA signaling is abnormal. In both hyperdopaminergic states, like drug abuse or schizophrenia (*27*), and hypodopaminergic states, like PD, the imposition of Hebbian rules on the sculpting of synaptic strength by experience will be disrupted, leading to inappropriate associations. This regulatory balance was absent in PD models. In this hypodopaminergic state, plasticity at glutamatergic synapses was still present, but it had lost its bidirectionality and dependence upon the temporal structure of pre- and postsynaptic activity. This might help to explain why striatal learning in PD patients is dysfunctional rather than simply absent (*28*). The cellular specificity of this imbalance also helps to explain the contrasting effects of DA depletion on the synaptic connectivity of D1 and D2 MSNs (*10*, *29*) and the longer-term adaptations in network activity thought to underlie the motor symptoms of PD (*30*).

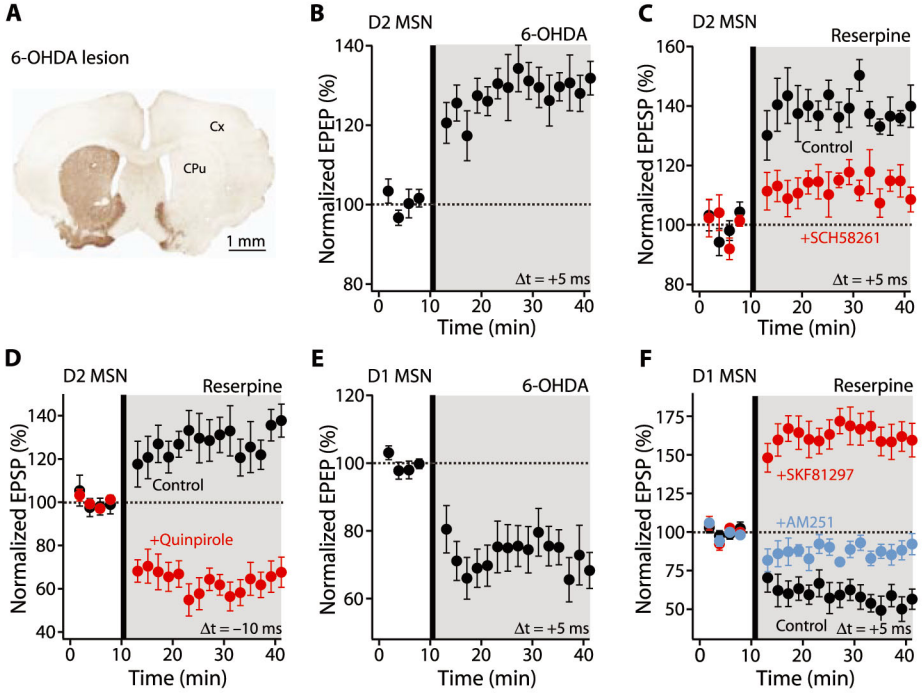


Fig. 3. Bidirectional Hebbian STDP is disrupted in MSNs from parkinsonian mice. (**A**) Light microscopic image of a coronal section showing the loss of immunoreactivity for TH after unilateral 6-OHDA lesioning. Cx, cortex; CPu, caudate-putamen. (**B** and **C**) LTP was induced from (B) lesioned D2 mice and (C) reserpine-treated mice after a positive timing protocol. Plot of average EPSP amplitude as a function of time. In (C), timing-dependent LTP induced in reserpine treated animals was blocked by SCH58261 (100 nM). (**D**) LTP also was induced with a negative timing protocol that would normally induce LTD. In contrast, perfusion of quinpirole (10 μM) restored LTD. (**E** and **F**) Timing-dependent LTD was evoked in D1 MSNs from (E) lesioned D1 mice and (F) reserpine-treated mice. In (F), D1 receptor agonist SKF81297 (3 μM) restored LTP after a positive timing protocol. The LTD induced in reserpine-treated mice was disrupted by AM-251 (2 μM).

References and Notes

1. A. M. Graybiel, T. Aosaki, A. W. Flaherty, M. Kimura, *Science* **265**, 1826 (1994).
2. H. H. Yin, B. J. Knowlton, *Nat. Rev. Neurosci.* **7**, 464 (2006).
3. J. N. Reynolds, B. I. Hyland, J. R. Wickens, *Nature* **413**, 67 (2001).
4. W. Schultz, *J. Neurophysiol.* **80**, 1 (1998).
5. D. J. Surmeier, J. Ding, M. Day, Z. Wang, W. Shen, *Trends Neurosci.* **30**, 228 (2007).
6. C. R. Gerfen *et al.*, *Science* **250**, 1429 (1990).
7. P. Calabresi, B. Picconi, A. Tozzi, M. Di Filippo, *Trends Neurosci.* **30**, 211 (2007).
8. A. C. Kreitzer, R. C. Malenka, *Nature* **445**, 643 (2007).
9. G. L. Gerdeman, J. Ronesi, D. M. Lovinger, *Nat. Neurosci.* **5**, 446 (2002).
10. M. Day *et al.*, *Nat. Neurosci.* **9**, 251 (2006).
11. W. Shen *et al.*, *Nat. Neurosci.* **10**, 1458 (2007).

12. P. J. Sjostrom, E. A. Rancz, A. Roth, M. Hausser, *Physiol. Rev.* **88**, 769 (2008).
13. T. Nevian, B. Sakmann, *J. Neurosci.* **26**, 11001 (2006).
14. V. Pawlak, J. N. Kerr, *J. Neurosci.* **28**, 2435 (2008).
15. Materials and methods are available as supporting material on *Science* Online.
16. Z. Wang *et al.*, *Neuron* **50**, 443 (2006).
17. L. Adermark, D. M. Lovinger, *J. Neurosci.* **27**, 6781 (2007).
18. M. A. Schwarzschild, L. Agnati, K. Fuxe, J. F. Chen, M. Morelli, *Trends Neurosci.* **29**, 647 (2006).
19. G. H. Seol *et al.*, *Neuron* **55**, 919 (2007).
20. T. Tzounopoulos, M. E. Rubio, J. E. Keen, L. O. Trussell, *Neuron* **54**, 291 (2007).
21. E. Fino, J. Glowinski, L. Venance, *J. Neurosci.* **25**, 11279 (2005).
22. J. P. Bolam, J. J. Hanley, P. A. Booth, M. D. Bevan, *J. Anat.* **196**, 527 (2000).
23. F. Gonon, *J. Neurosci.* **17**, 5972 (1997).
24. E. K. Richfield, J. B. Penney, A. B. Young, *Neuroscience* **30**, 767 (1989).
25. F. A. Middleton, P. L. Strick, *Brain Res. Brain Res. Rev.* **31**, 236 (2000).
26. K. Nakamura, O. Hikosaka, *J. Neurosci.* **26**, 5360 (2006).
27. P. R. Montague, S. E. Hyman, J. D. Cohen, *Nature* **431**, 760 (2004).
28. K. Dujardin, B. Laurent, *Curr. Opin. Neurol.* **16** (suppl. 2), S11 (2003).
29. N. Mallet, B. Ballion, C. Le Moine, F. Gonon, *J. Neurosci.* **26**, 3875 (2006).
30. M. D. Bevan, P. J. Magill, D. Terman, J. P. Bolam, C. J. Wilson, *Trends Neurosci.* **25**, 525 (2002).
31. This work was supported by NIH grants NS 34696 (D.J.S.), MH074866 (P.G., D.J.S.), and DA10044 (P.G.) and a grant from the Picower Foundation (D.J.S., P.G.). We thank Q. Ruan, K. Saporito, and S. Ulrich for their technical assistance; E. Ilijic for generating brain images of tyrosine hydroxylase immunoreactivity following unilateral 6-OHDA lesioning; M. Day for supplying two-photon laser scanning microscopy images; N. Spruston, R. Malenka, and the members of the Surmeier laboratory for helpful discussion; and N. Spruston and A. Contractor for their comments on the manuscript.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5890/848/DC1
Materials and Methods
Figs. S1 to S7
References

16 May 2008; accepted 4 July 2008
10.1126/science.1160575

Dynamic Shifts of Limited Working Memory Resources in Human Vision

Paul M. Bays[1,2]* and Masud Husain[1,2]

Our ability to remember what we have seen is very limited. Most current views characterize this limit as a fixed number of items—only four objects—that can be held in visual working memory. We show that visual memory capacity is not fixed by the number of objects, but rather is a limited resource that is shared out dynamically between all items in the visual scene. This resource can be shifted flexibly between objects, with allocation biased by selective attention and toward targets of upcoming eye movements. The proportion of resources allocated to each item determines the precision with which it is remembered, a relation that we show is governed by a simple power law, allowing quantitative estimates of resource distribution in a scene.

The dominant model of visual memory capacity asserts that only a limited number of items can be simultaneously represented in working memory (*1*–*10*). Support for this model has come primarily from change-detection tasks, in which detection was close to 100% correct for small numbers of items, and declined only when the set size increased above a certain limit, generally three or four items (*4*, *5*). An alternative way to explore the limits of visual working memory is to consider the precision with which each visual item is stored as a function of the number of objects in a scene. This approach provides a radically different perspective on visual capacity limits, revealing rapid redistribution of limited memory resources across eye movements and covert shifts of attention.

We tested subjects' ability to remember the location and orientation of multiple visual items after a brief disappearance of the stimulus array, with or without an intervening eye movement (Fig. 1). To minimize the role of configurational memory (*11*), only one of the items was redisplayed after the delay; subjects reported the direction in which this probe item had been displaced or rotated. Responses varied probabilistically with the magnitude of the change to the probe item (Fig. 2A). Subjects' response functions were successfully fitted with cumulative Gaussian distributions, consistent with a Gaussian distribution of error in the stored representation of the original stimulus (*12*).

In the absence of eye movements, subjects were able to recall both location and orientation of a single item with considerable accuracy (Fig. 2A, $N = 1$, black symbols), with discrimination of 0.5° displacements and 5° rotations significantly better than chance at 73% and 80% correct, respectively ($t > 5.8$, $p < 0.001$). However, increasing the number of items to be remembered led to a decrease in performance, indicative of the limited capacity of visual working memory (Fig. 2A, black symbols, set size increasing left to right). Precision, measured by the reciprocal of the standard deviation of the response function, was reduced as the number of items in the display increased (Fig. 2B, black symbols). These data do not reveal a sharp drop in performance at a limit of four items.

Next, we asked whether the precision of visual working memory is affected by an eye movement. Detection of changes to visual stimuli that occur during an eye movement presents a challenge to the brain, because the pre- and post-saccadic retinal locations of every visual item are very different. For location discrimination in single-item displays (Fig. 2A, top left, red symbols), an intervening eye movement introduced a small bias (mean 1.4°) into subjects' judgments: a tendency to report a shift in the direction of the saccade even for small displacements in the opposite direction. However, as can be seen from the similar slopes of the two response functions, the precision with which this discrimination was made did not differ significantly from that of the fixation condition ($t = 1.2$, $p = 0.24$). This indicates that subjects take into account the size and direction of their eye movement in order to esti-

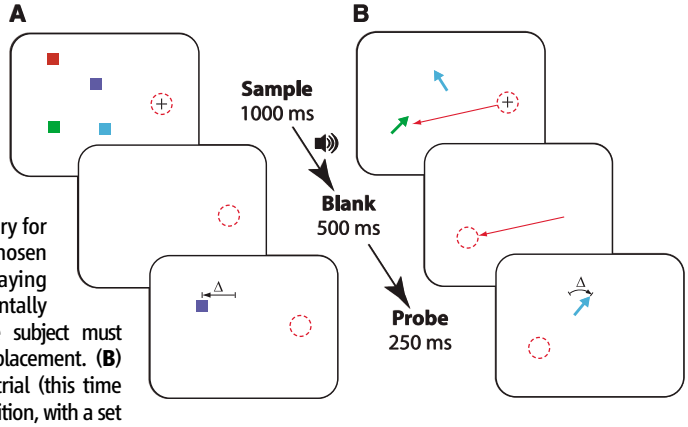


Fig. 1. Experimental procedure. (**A**) Stimuli and sequence of events on a location-judgment trial. The example shown is a fixation trial with a set size of four items. After the sample display is blanked, subjects' memory for location of a randomly chosen item is tested by redisplaying the item displaced horizontally through distance Δ. The subject must report the direction of displacement. (**B**) An orientation judgment trial (this time shown for the saccade condition, with a set size of two items). At the tone, the subject makes a saccade toward the item of a prespecified color (here green) with the display being blanked during the eye movement. A randomly chosen item is redisplayed, rotated through an angle Δ, and the subject reports the direction of rotation. Red circles indicate gaze position.

[1]Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR, UK. [2]Institute of Neurology, University College London, Queen Square, London WC1N 3BG, UK.

*To whom correspondence should be addressed. E-mail: p.bays@ion.ucl.ac.uk

mate the expected postsaccadic retinal location of the single target (*13*, *14*). This may be achieved by remapping a retinotopic location representation based on an internal copy of the saccadic motor signal (*15*, *16*).

Precision in the saccade condition decreased with increasing number of items in a way similar to that of the fixation condition, for both location and orientation judgments (Fig. 2B, red symbols), with no significant advantage of fixation at any of the tested set sizes ($t < 1.3$, $p > 0.23$). This indicates that the process of spatial updating does not introduce any additional capacity limit on visual working memory, and therefore that the full contents of memory undergo remapping.

The item-limit model of visual working memory predicts that discrimination performance will begin to decline only once the limiting number of items is exceeded. In contrast, our results—for both fixation and saccade conditions—show that the precision with which visual items are remembered decreases with increasing numbers even at the smallest set sizes ($t > 2.7$, $p < 0.006$), with the largest drop in precision occurring between one- and two-item displays, and no evidence for any discontinuity in the region of four items (Fig. 2B). Our data therefore support an alternative model in which limited visual memory resources must be shared out between items, such that increasing numbers of items are stored with decreasing precision (see fig. S1 for an illustration). To quantify the relation between the resources available to encode an item (R) and the precision with which it is remembered (P), we replotted precision as a function of the proportion of resources available per item (Fig. 3A). The results suggest that this relation can be captured by a simple power law [$P \propto R^k$, power constant $k = 0.74 \pm 0.06$ (95% confidence limits); blue line].

The similarity of our results for memory of location and orientation suggests that they share a common mechanism. This may be the representation of stimulus attributes by population coding, in which information is encoded in the combined activity of a large number of neurons (*17*). Currently identified population decoding schemes do not permit a neuron to simultaneously encode information about more than one stimulus. Therefore, when multiple items must be represented, the total pool of neurons must be shared out between the different items. Because each neuron's firing rate is corrupted by noise (*18*), reducing the number of neurons representing an item will increase variability in the population estimate, and consequently reduce the precision with which the item is represented. Theoretical studies have shown that a maximum likelihood decoding scheme would result in a power-law relation between precision and number of neurons (*19*), similar to that obtained in the current study (*20*).

Can the power-law model also explain why previous studies (*4*, *5*) found a decrease in performance only for greater numbers of items? Figure 3B shows how the model predicts precision will change with increasing set size, and Fig. 3C displays the corresponding response functions. The power-law model predicts that accuracy (proportion correct) will vary with the magnitude of the change to be discriminated. In this study, we tested discrimination of small changes to stimuli, where discrimination is difficult even with only one item in the display. In this range, our model predicts that accuracy will decrease with increasing number of items even at the smallest set sizes (e.g., dotted vertical line in Fig. 3C). In contrast, previous tests of visual working memory have generally used "suprathreshold" changes, where performance is close to 100% correct for a single item. In these cases, the power-law model predicts that accuracy will initially change almost imperceptibly with increasing numbers of items, and then more steeply at larger set sizes (e.g., dashed vertical line in Fig. 3C). The full predictions of the model are shown in Fig. 3D (black lines). The power-law model is consistent both with our data (examples shown in red: 0.5° displacements, 5° rotations; see also fig. S2) and with many of the results previously taken to support a three- to four-item limit [examples shown in green (*5*)].

Although we have shown that an upcoming eye movement does not reduce the total memory resources available, it does affect how those resources are allocated. Figure 4A shows precision of discrimination judgments in multi-item ($N > 1$) displays, where the data were separated into trials on which the probed item was the saccade target and those on which it was one of the other items in the display. For both location and orientation judgments, the saccade target was remembered with greater precision than were nontargets, indicating a preferential shift of visual memory resources to the target of the eye movement (black symbols; $t > 4.2$, $p < 0.001$). This finding was not a consequence of the way in which the saccade target was specified [endogenously cued by color (*21*)] because a similar effect was also observed, in a different condition, when we cued

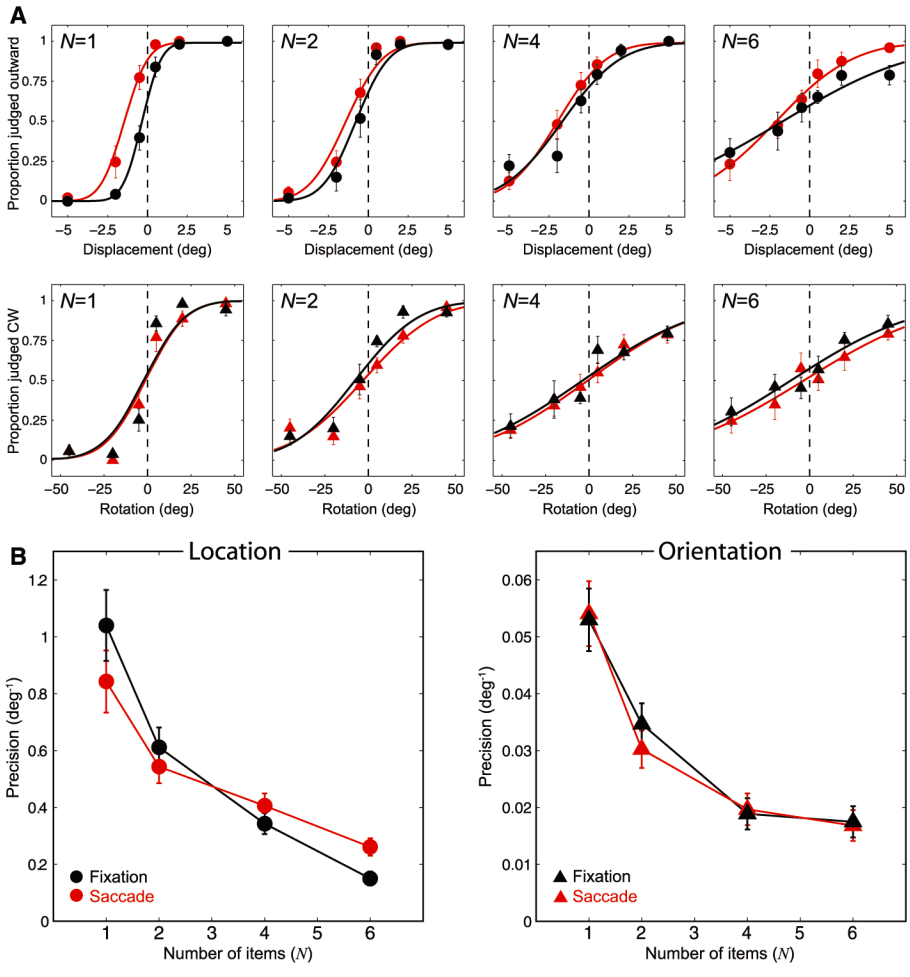


Fig. 2. Performance on the memory task. (**A**) Proportion of displacements judged outward from fixation (top row) as a function of the actual displacement (Δ) and number of items in the display (*N*). Similar plots are shown in the lower row for orientations (rotations) judged clockwise as a function of the actual rotation and number of items. Results from fixation trials are shown in black, and those from saccade trials in red. Curves indicate cumulative Gaussian distributions fitted to the response data. The slopes of these functions become flatter with increasing number of items. (**B**) Precision (determined by the reciprocal of the SD of the fitted Gaussian) falls as a function of the number of items in the sample display. Error bars indicate ±1 SE.

the saccade target exogenously by flashing it (Fig. 4A, gray symbols; $t > 4.3$, $p < 0.001$). Thus, limited working memory resources get rapidly redistributed so that the target of a forthcoming eye movement receives privileged allocation, thereby improving the precision for this item. Because total resources are limited, the corollary of this enhanced memory for the saccade target should be a decrease in precision for nontargets, which will be most evident when the total number of items is small. A comparison of saccade and fixation performance in two-item displays confirmed this effect, with the increased precision for the saccade target ($t = 4.26$, $p < 0.001$) matched by a significant decrease in precision for the nontarget item ($t = 3.19$, $p < 0.01$).

Is this flexibility in the allocation of memory resources specific to eye movements, or does it also occur with shifts of covert attention (*22*–*25*)? In a further condition, subjects kept their eyes fixed, but one of the items in the sample display flashed briefly before the screen was blanked, a manipulation known to involuntarily attract visual attention (*26*). When the flashed item was subsequently probed, discrimination precision was significantly higher than for nonflashed items [Fig. 4B; $t > 3.4$, $p < 0.001$ (*27*)]. Thus, visual attention acts as a "gatekeeper," determining which visual information is given priority for storage in working memory (*28*–*31*), perhaps by biasing competitive interactions in cortical regions mediating visual memory (*32*, *33*).

In normal scene viewing, we make many eye movements in order to extract the maximum possible information from a scene. We performed an additional experiment to examine how visual memory resources are dynamically allocated across a sequence of saccades. Subjects made a series of eye movements to fixate each item in a five-item display; the display was blanked before the saccade to the fifth item reached its target. The precision on a subsequent discrimination judgment, probing memory for any one of the five targets, varied with order of fixation (Fig. 4C).

Discrimination of both location and orientation of the saccade target (the fifth item) was substantially more precise than for any of the other items in the display ($t > 3.9$, $p_{corrected} < 0.001$). However, the target of the previous saccade, which was also the most recently fixated item, was not remembered with significantly greater precision than any of the previously fixated items ($t < 2.5$, $p_{corrected} > 0.12$). Nor were there any differences in precision between the previous items ($t < 2.0$, $p_{corrected} > 0.55$). We found no significant relation between precision and fixation time ($t < 1.0$, $p > 0.31$), indicating that these results do not reflect temporal (e.g., recency) effects. Rather, it appears that the high-resolution memory for a saccade target persists for only one eye movement. Based on the power law obtained in the first experiment, we can estimate the proportion of working memory resources allocated to each item in the sequence. This analysis reveals that, at the time of a saccade, most visual memory resources are allocated to the target of the next fixation [location task: 56%; orientation task: 61% (*34*)] rather than to the currently fixated item (location task: 15%; orientation task: 16%).

The current results are inconsistent with the view that visual working memory capacity is limited to a fixed number of objects. Several previous studies have attempted to go beyond the simple fixed item-limit account of visual memory (*9*, *10*, *35*). One study (*9*) has proposed a variable item-limit, based on a fixed "information load," whereby the more visually complex the items to be remembered, the fewer can be stored. Although

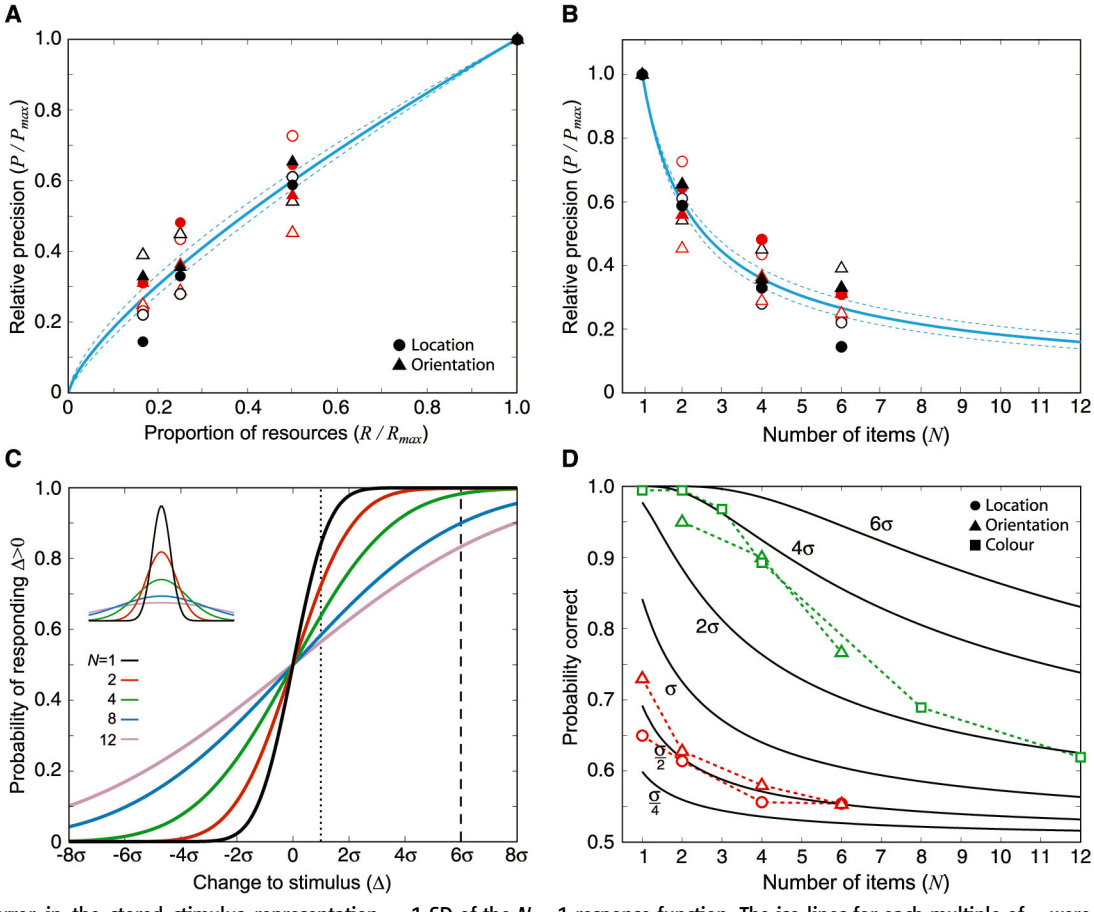


Fig. 3. Modeling visual memory performance. (**A**) The relation between available memory resources and precision is approximated by a power law (solid blue line; dashed line indicates 95% confidence limits) fitted to the normalized precision values obtained in the first experiment, including both fixation and saccade conditions (circles: location task; triangles: orientation task; black: fixation trials; red: saccade trials; empty symbols: flash cue; filled symbols: no flash cue). (**B**) Normalized precision as a function of number of items in memory (N) in all conditions. Solid line indicates the prediction of the fitted power-law model. Normalization is with respect to performance with one item ($N = 1$) in each of the experimental conditions. (**C**) Response probability as a function of the size of the change (Δ) to the stimulus (i.e., displacement or rotation), for different numbers of items (N), predicted on the basis of the power-law model. σ indicates 1 SD of the $N = 1$ response function. The curves become flatter with increasing number of items, corresponding to changes in the Gaussian distributions of error in the stored stimulus representation (inset). The dotted vertical line corresponds to a small change to the stimulus, as used in the current study, whereas the dashed vertical line indicates a much larger change. In the latter case, performance would be near-maximal for one to four items but fall with further increases in set size. (**D**) Probability of correct response for stimulus changes of different magnitudes (black lines). σ indicates 1 SD of the $N = 1$ response function. The iso-lines for each multiple of σ were derived directly from (C). Red symbols show empirical data from the current study. Green symbols show data from (*5*). Both sets of data are consistent with the power-law model, but different curves arise with differences in the size of stimulus change. See also fig. S2 for data plotted according to different sizes of change within the current experiment.

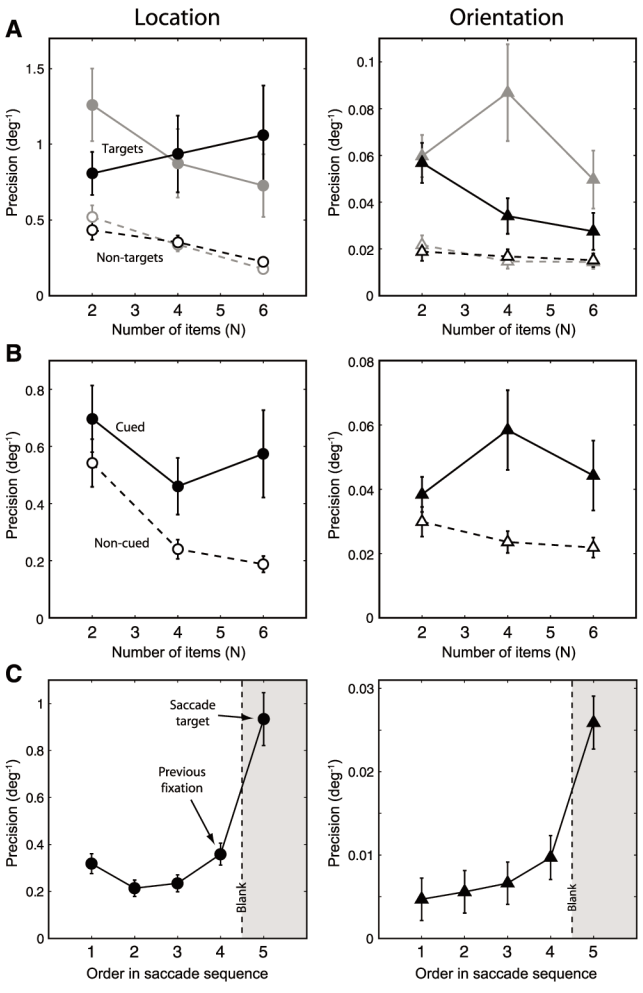


Fig. 4. Effects of eye movements and attention. (**A**) Precision as a function of number of items: memory for saccade targets (filled symbols) and nontargets (empty symbols), when the target is specified endogenously by color (black) or exogenously by a flash (gray). Better performance is seen for targets, regardless of the mode of cueing. (**B**) Memory for an item cued by a flash (filled symbols) and for noncued items (empty symbols), with no eye movements. (**C**) Memory for items as a function of fixation order in a sequence of saccades demonstrates how memory for the most recently fixated item is poor compared to memory for the current saccade target. Error bars indicate ±1 SE.

related to our limited-resource model, this hypothesis cannot account for the relation between precision and number of items observed in the current study, because the visual complexity of the sample stimuli was held constant. It has been argued (*10*) that the changes in detection performance observed in this previous study are the result of increasing similarity between sample and probe items, rather than increasing complexity of the sample. Because the precision of visual memory is limited, reducing the size of the change to the stimulus results in poorer performance, in agreement with our model.

Since submission of this article, another study has been published that also examines the precision of visual memory (*36*). The authors put forward a two-component model, combining a variable-precision memory for fewer items with an absolute upper limit on number of items (above which decreases in performance are accounted for solely by random guesses). Based on this interpretation, their data indicate that the average subject can hold only about two items in working memory [see figure 2 and supplementary figure 3 in (*36*)]. However, this study did not control eye movements, which we have shown can strongly bias precision in favor of fixation targets. A reanalysis of our own fixation task data in accordance with their mixture-model approach reveals that precision falls with increasing number of items throughout the tested range, including between four and six items ($\chi^2 = 5.6$, $p = 0.018$; fig. S3 and supporting online text). We conclude, therefore, that the capacity of visual memory can be explained solely in terms of a limited resource that must be shared out between all items in the scene, with no evidence for an upper limit on the number of items that can be stored, contrary to the hypothesis of a two-component model (*36*).

The allocation of this limited resource is highly flexible; making an eye movement to an item, or directing covert attention to it, causes a greater proportion of memory resources to be allocated to it, so it is retained with far greater precision than other objects in the scene. All information stored in visual working memory is dynamically updated during an eye movement to take into account the change in gaze position. However, because the resource is limited, the high-resolution representation of a fixated item is substantially degraded as memory resources are reallocated to the target of the next eye movement.

References and Notes

1. G. Sperling, *Psychol. Monogr.* **74**, 1 (1960).
2. H. Pashler, *Percept. Psychophys.* **44**, 369 (1988).
3. D. E. Irwin, *J. Exp. Psychol. Learn. Mem. Cogn.* **18**, 307 (1992).
4. S. J. Luck, E. K. Vogel, *Nature* **390**, 279 (1997).
5. E. K. Vogel, G. F. Woodman, S. J. Luck, *J. Exp. Psychol. Hum. Percept. Perform.* **27**, 92 (2001).
6. N. Cowan, *Behav. Brain Sci.* **24**, 87 (2001).
7. E. K. Vogel, M. G. Machizawa, *Nature* **428**, 748 (2004).
8. J. J. Todd, R. Marois, *Nature* **428**, 751 (2004).
9. G. A. Alvarez, P. Cavanagh, *Psychol. Sci.* **15**, 106 (2004).
10. E. Awh, B. Barton, E. K. Vogel, *Psychol. Sci.* **18**, 622 (2007).
11. Y. Jiang, I. R. Olson, M. M. Chun, *J. Exp. Psychol. Learn. Mem. Cogn.* **26**, 683 (2000).
12. T. D. Wickens, *Elementary Signal Detection Theory* (Oxford Univ. Press, New York, 2002).
13. H. Deubel, W. X. Schneider, B. Bridgeman, *Vision Res.* **36**, 985 (1996).
14. M. L. Davidson, M. J. Fox, A. O. Dick, *Percept. Psychophys.* **14**, 110 (1973).
15. P. M. Bays, M. Husain, *Neuroreport* **18**, 1207 (2007).
16. J. R. Duhamel, C. L. Colby, M. E. Goldberg, *Science* **255**, 90 (1992).
17. A. Pouget, P. Dayan, R. Zemel, *Nat. Rev. Neurosci.* **1**, 125 (2000).
18. W. Bialek, F. Rieke, *Trends Neurosci.* **15**, 428 (1992).
19. H. S. Seung, H. Sompolinsky, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10749 (1993).
20. Simple models of maximum likelihood decoding predict a power constant of 0.5, significantly smaller than that obtained empirically in the current study (0.74). This discrepancy may be a result of the simplifying assumption that neurons in the population fire independently of one another; see, e.g., (*37*).
21. Materials and methods are available on *Science* Online.
22. C. W. Eriksen, J. E. Hoffman, *Percept. Psychophys.* **14**, 155 (1973).
23. M. I. Posner, *Q. J. Exp. Psychol.* **32**, 3 (1980).
24. H. Deubel, W. X. Schneider, *Vision Res.* **36**, 1827 (1996).
25. J. E. Hoffman, B. Subramaniam, *Percept. Psychophys.* **57**, 787 (1995).
26. S. Yantis, J. Jonides, *J. Exp. Psychol. Hum. Percept. Perform.* **10**, 601 (1984).
27. As in the saccade condition, this reallocation of resources resulted in both an increase in mean precision for the flashed item and a decrease in mean precision for nonflashed items, although in this case these differences did not attain statistical significance.
28. R. Desimone, J. Duncan, *Annu. Rev. Neurosci.* **18**, 193 (1995).
29. N. Cowan, *Attention and Memory: An Integrated Framework* (Oxford Univ. Press, New York, 1995).
30. E. Awh, E. K. Vogel, S.-H Oh, *Neuroscience* **139**, 201 (2006).
31. A. Treisman, *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **353**, 1295 (1998).
32. R. Desimone, *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **353**, 1245 (1998).
33. J. Duncan, G. Humphreys, R. Ward, *Curr. Opin. Neurobiol.* **7**, 255 (1997).
34. We do not claim that the fraction of resources allocated to the saccade target is fixed. Rather, this allocation may be flexible depending on factors such as the number of nontarget items and specific task demands.
35. J. Palmer, *J. Exp. Psychol. Hum. Percept. Perform.* **16**, 332 (1990).
36. W. Zhang, S. J. Luck, *Nature* **453**, 233 (2008).
37. B. B. Averbeck, P. E. Latham, A. Pouget, *Nat. Rev. Neurosci.* **7**, 358 (2006).

38. We thank M. Bays, A. Faisal, M. Machizawa, K. Nzerem, and P. Sumner for helpful discussion. This research was supported by the Wellcome Trust and the National Institute for Health Research Clinical Biomedical Centre at University College London Hospitals/University College London.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5890/851/DC1
Materials and Methods
Figs. S1 to S3
References

19 March 2008; accepted 4 July 2008
10.1126/science.1158023

New Products

Automated XY Stage

The MS-9400 XY stage is designed to provide a high resolution, highly repeatable means of controlling the X and Y positions of the microscope stage. The unit offers a travel range of 225 mm by 100 mm to allow it to scan eight 25-mm by 75-mm slides. All axes derive their precise control through the use of closed-loop DC servometers employing high-resolution encoders for positioning feedback. The XY stage features crossed-roller slides, high-precision lead screws, and zero-backlash miniature-geared DC servometers to provide 100-nm resolution with repeatability accuracy of less than 700 nm rms (root mean square) using standard rotary encoders. Optional linear encoders can achieve resolutions of 50 nm and repeatability accuracies of less than 300 nm rms.

Applied Scientific Instrumentation
For information 541-461-8181
www.ASIimaging.com

High Performance PCR

The High Performance PCR (polymerase chain reaction) system combines highly processive proofreading Phusion DNA Polymerases, high-speed Piko Thermal Cyclers, and ultrathin-walled tubes and plates. Compared with conventional PCR, the system provides faster protocols with accurate and specific amplification yet excellent yields. It offers superior accuracy compared with Taq- and Pfu-based systems and reduced levels of primer-dimers and false-primed products. The system saves protocol time, bench space, reaction volume, and per-run cost.

Finnzymes
For information +358-9-2472-3291
www.finnzymes.com

Desalting Tubes

Clarity Desalting Tubes are for basic reverse-phase removal of salt and excess reagents from synthetic oligonucleotides. Delivering a purity level of about 70 percent with 80 percent target oligonucleotide recovery, the tubes offer a faster and more cost-effective solution than traditional gel filtration methods. For higher throughput, the tubes can be multiplexed using a vacuum manifold.

Phenomenex
For information 310-212-0555
www.phenomenex.com

Liquid Flow Management

The VROCT ("Viscometer/Rheometer-on-a-Chip" technology) enables microliter-size samples to be used for high resolution, accurate, and repeatable viscosity measurement. The small chip geometry extends viscosity measurements beyond conventional limits into the realm of very high shear rates, at the conditions occurring in actual applications.

Rheosense
For information 925-866-3802
www.RheoSense.com

Freezer Labeling System

The Brady Freezerbondz labeling system ensures secure labeling in low temperature storage conditions. Freezerbondz II is a polyester material with a permanent adhesive that can be applied directly to already frozen surfaces without the need to overlap, eliminating the risk attached to thawing samples for labeling. These labels remain secure even when samples are transferred from liquid nitrogen at -196°C to a water bath at 100°C or even to an autoclave. When printed with the correct Brady Thermal Transfer print system, these labels withstand solvents such as ethanol toluene, xylene, and DMSO. The labels are available for vials, slides, and well plates, and in custom sizes.

Brady Corp
For information 01295-228288
www.bradylab.co.uk

Laser Scanning Microscope

The LSM 710 Laser Scanning Microscope provides new possibilities in research conducted with living, multilabeled cells. With its sensitivity, high signal-to-noise ratio, flexibility for new fluorescent dyes and multimodal experiments as well as new multiphoton detectors for deeper optical penetration into biological structures, the system can give new impetus to all areas of biological research. The system also features unique precision and reproducibility and easy operation. It offers high-contrast, detailed images of even the most complex specimens, such as thick, living tissue samples. The new illumination and detection design provides freedom in the selection and simultaneous imaging of up to 10 fluorescent signals. The basis is a filter-free spectral detection unit, which can be continuously set over the entire wavelength range. Innovative analysis methods such as integrated image correlation spectroscopy make it possible to extract quantitative information about molecule concentrations and mobility directly from the confocal images recorded.

Carl Ziess
For information 800-233-2343
www.zeiss.de/sound

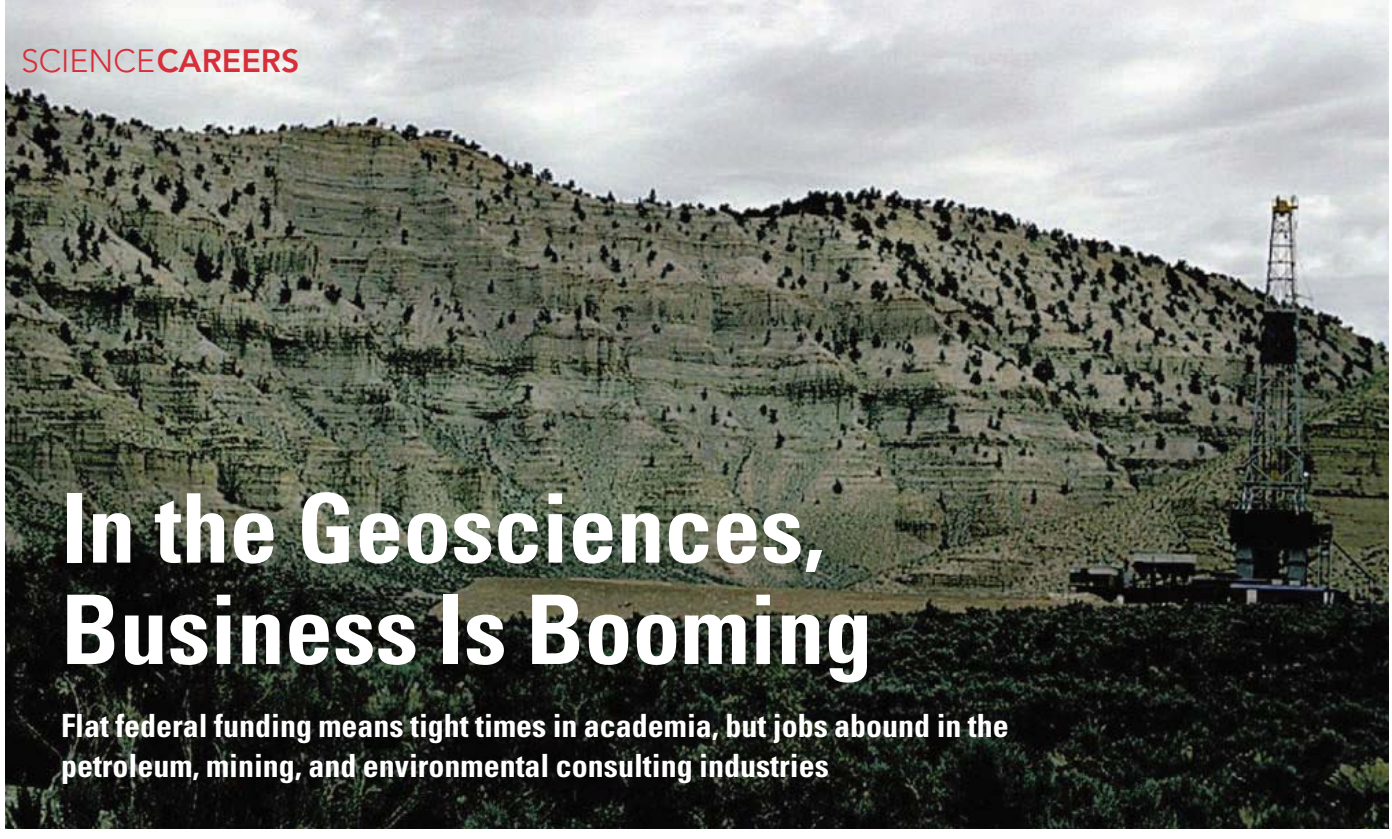
Protease Inhibitor Cocktail

ProteaseArrest is a general protease inhibitor cocktail that is provided as a 100X concentrated solution that inhibits more than 95 percent of protease activity, even in tissues known to have higher than normal protease activity. ProteaseArrest is suitable for the protection of proteins during preparation of samples and protein purification from animal tissues, plants, yeast, and bacteria. For added convenience, it is available in single-use aliquots capable of greater than 95 percent protease inhibition in 10-ml solutions. Other formats are available that require no dissolving or resuspension.

G-Biosciences/Genotech
For information 314-991-6034
www.GBiosciences.com

Electronically submit your new product description or product literature information! Go to **www.sciencemag.org/products/newproducts.dtl** for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availabilty of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.



In the Geosciences, Business Is Booming

Flat federal funding means tight times in academia, but jobs abound in the petroleum, mining, and environmental consulting industries

GEOLOGIST AMY SIMONSON LOVES HER work. She spends her days in the countryside around Charlottesville, Virginia, measuring stream flow and groundwater levels for the state's Department of Environmental Quality. The job, she says, is exactly what she wanted.

Simonson, 25, began her job hunt in 2007 after getting a master's degree in geology from the University of Delaware, Newark. She had one condition: She wanted to spend as much time as possible in the field, not in front of a microscope or a computer. Taking a scattershot approach, she applied for jobs in geophysics, engineering, environmental consulting, and geographic information system mapping. She didn't have to wait long. "I got offered a lot of stuff," she says.

Online
sciencemag.org
Hear more about geology careers in this week's *Science* podcast.

Simonson's experience isn't rare. For many young geoscientists now embarking on careers, the job outlook is very good. The current federal research funding situation means it's less rosy for those on an academic research track. But for those in industry, the number of geoscience jobs will grow by 22% from 2006 to 2016, much faster than the projected total of a 10% increase for all occupations, according to the U.S. Bureau of Labor Statistics.

"In general, the market is hot," says Cindy Martinez, who analyzes geoscience workforce issues at the American Geological Institute (AGI) in Alexandria, Virginia. "Functionally, there's no unemployment of geoscientists right now."

In the petroleum, mining, and environmental consulting industries, a desperate quest for new talent has sent companies scrambling to hire new graduates. Traditionally, a master's has been the professional degree of choice for industry employers. But the need for new hires within these fields is such that even graduates with bachelor's degrees are finding jobs, particularly in environmental consulting—although a master's is generally needed to move up the ladder from fieldwork to the office.

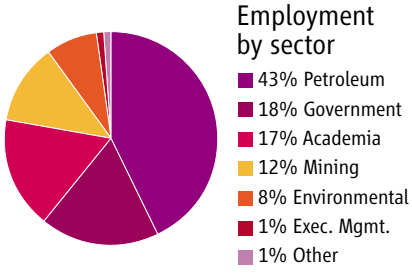
That intense competition for new hires has raised starting salaries in these industries, especially oil: Graduates now entering the petroleum industry earn $82,500 a year, on average, according to AGI.

Geoscience salaries generally have also been increasing, AGI data show. In 2005, the average starting salary for a geoscientist in an industry, academic, or government position was $74,000, a 9.7% increase over 2004. For later career scientists with more than 20 years of experience, the average salary was $139,000 in 2005, an increase of more than 23% over the previous year.

The current hiring boom in the petroleum industry is a welcome change from the layoffs of the 1980s and 1990s; unemployment among geoscientists reached 11% in 1985. Those layoffs left a distinct gap in the oil industry workforce between new hires and senior managers, a 2007 National Petroleum Council report noted. And with many senior managers likely to retire within the next decade, there aren't enough midlevel managers ready to take the helm. This hiring and firing pattern is "totally cyclical," Martinez says. "The industry needs to work on fixing that."

Although industry jobs are readily available, the job market is tougher for students seeking careers in academia. One problem is a research-funding shortage, as primary funding sources such as the National Science

GEOSCIENCE STATS



Starting Salaries

	AVERAGE ($)	MEDIAN ($)
Geoscience Masters		
Oil and Gas	81,300	82,500
Environmental Firm	47,500	45,500
Any Government	46,200	45,000
Geoscience Ph.D.s		
Postdoc—Academe	43,100	42,000
Postdoc—Gov't.	55,200	53,000
Permanent Academe	51,900	52,500
Private Sector	72,600	71,000

PHOTO CREDIT: BRUCE MOLNIA/TERRA PHOTOGRAPHICS; GRAPH SOURCE: NSF/AGI/BLS

Foundation have suffered from essentially flat research budgets for the past few years. That lack of research money contributes to another trend: When a venerated geology professor retires, some universities are choosing not to hire a replacement, filling the position with a scientist in a different, often environmental, field, or not at all. As of January 2008, the number of geoscience faculty members in U.S. colleges and universities was 12,354, down from 13,554 in 1999, according to a June 2008 report released by AGI.

The lack of available academic positions has left geoscientists who aspire to faculty jobs in limbo, often stringing together several postdoc positions as they wait for openings, Martinez says. The percentage of geoscience postdocs "has really gone through the roof," with about 58% of Ph.D. graduates pursuing postdocs in 2005 compared with 40% in 1999. That trend may soon slow if a combination of strong industry salaries and weak academic opportunities pushes some students to leave graduate school early. "We're seeing geoscientists in droves going into industry with master's degrees and not staying on for Ph.D.s," Martinez says.

"My perception is indeed that there are many more applicants for jobs than there are positions," says Joseph Colgan, a Mendenhall Postdoctoral Research Fellow at the U.S. Geological Survey (USGS) in Menlo Park, California. Colgan, who studies the geologic setting of mineral deposits in the western United States, has considered academic jobs but would like to stay in a more permanent job at USGS. However, the agency, like many universities, has tightened its belt after years of flat or declining budgets, which means fewer hires.

Geoscience graduates will have training that qualifies them for jobs outside of their home field. Scientists with training in multidisciplinary specialties such as isotopic tracers, mineral commodities, and geotechnology are successfully venturing into fields that aren't considered geologic, including medicine, law, and finance. In fact, only 50% of people with geoscience degrees currently work in the geosciences, according to AGI.

Ultimately, for geoscientists who want to work in industry, job opportunities abound. "The earth sciences are in a somewhat unique situation at the moment because we're in one of the biggest commodity booms ever," Colgan says. So if he doesn't get a permanent position with USGS, he adds, "I'll come up with something else."

–CAROLYN GRAMLING

Carolyn Gramling is a geosciences writer in Washington, D.C.

CREDIT: SCHLUMBERGER

In the field. Sally Serenyi joined Schlumberger for "a bit of adventure."

GEOSCIENTISTS IN HIGH DEMAND IN THE OIL INDUSTRY

The next generation of petroleum geologists will face unique challenges in meeting the world's energy demands

SIX YEARS AGO, WOULD-BE LAWYER KIRA DIAZ-TUSHMAN HEARD A NATIONAL PUBLIC Radio program about the impending retirement of senior U.S. Geological Survey (USGS) geologists. "I thought, 'That sounds fun. I want to do what they're doing and play around in the field.' " So she double-majored in geology and political science at Bryn Mawr College in Pennsylvania and did a summer internship at USGS.

Watching researchers scramble for dwindling federal funding turned her away from government work. But her interest in geology persisted, so she studied for a master's degree in structural geology at the University of Texas (UT), Austin. While there, she interned with the technology group at Apache Corp., an oil exploration company based in Houston, and learned the basics of seismic mapping. She began to seriously consider a job in oil exploration.

Diaz-Tushman, now an operations geologist for BP, is part of a fast-growing global cadre of scientists and engineers building careers around unlocking more of Earth's energy reserves. Those in the field repeat the mantra that the "easy oil is gone"; this new generation faces the challenge of finding oil in remote locations and of pioneering new ways to tap into unconventional reserves in existing oil fields.

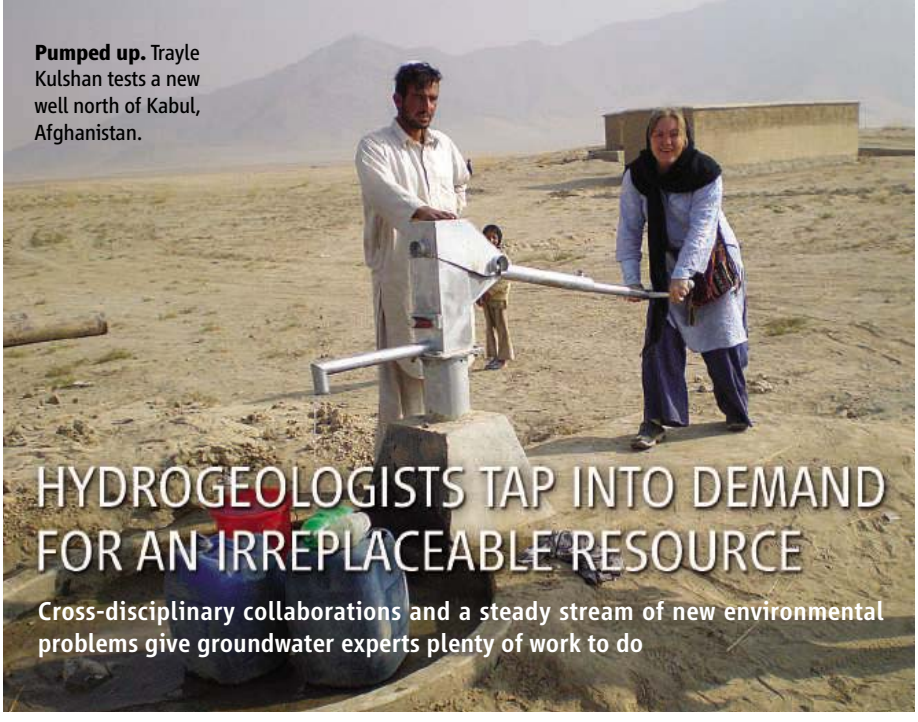
Mind the gap

The cyclical nature of oil prices has left a historical footprint on the existing pool of human resources. Low prices in the 1980s and 1990s meant that many major companies recruited less or not at all, leaving them top-heavy and in need of young talent.

The hiring gap "is more of a driver than the price" of oil today, says Laura DeMott, a Houston-based senior petroleum geologist at ExxonMobil. Regardless of the cause, demand for people with geological and geophysical training is high in the oil industry, and experts predict it will stay strong for the next 5 to 10 years.

People entering the industry will have their choice of a great diversity of companies, locations, and career paths. In North America, family-owned single-drill outfits work alongside multinational

continued on page 859



Pumped up. Trayle Kulshan tests a new well north of Kabul, Afghanistan.

HYDROGEOLOGISTS TAP INTO DEMAND FOR AN IRREPLACEABLE RESOURCE

Cross-disciplinary collaborations and a steady stream of new environmental problems give groundwater experts plenty of work to do

WHEN HYDROGEOLOGISTS TALK ABOUT their field, one word keeps coming up: "recession-proof." While geologists in the energy and mineral industries face roller-coaster hiring-and-firing cycles, those who study the movement and chemistry of water seeping through rocks and sediment find demand for their expertise almost as steady as the flow of groundwater itself.

"I can't think of any unemployed hydrogeologists," says Roy Haggerty, an associate professor of hydrogeology at Oregon State University, Corvallis. It's easy to see why. Water is essential, irreplaceable, and, as populations and economies grow, increasingly in demand and endangered.

Environmental consulting companies, which employ about 80% of hydrogeologists in the United States, currently report four jobs for every qualified graduate, according to the American Geological Institute (AGI). Government regulatory agencies, national laboratories, and mining and oil companies also need them. New niches open regularly as hydrogeologists team up with scientists in other disciplines to grapple with huge environmental challenges, such as forecasting how changing climate will affect water resources and aquatic life. And signs are that the future will hold more of the same. As Richelle Allen-King, a hydrogeology professor at the University at Buffalo in New York, puts it, "Water problems are not going away."

A steady flow

Kurt Zeiler, 32, works in the Denver, Colorado, office of the global environmental-services company AMEC Geomatrix. After 5 years as a hydrogeologist, he says: "It's going well. I definitely love this field. There's a lot of opportunities to do really interesting science."

Zeiler's training—a B.S. in earth science from Montana State University in Bozeman and a double-M.S. in water resources and hydrogeology from the University of Wisconsin, Madison—exposed him to the whole gamut of hydrogeologic work. He learned to monitor the water levels in wells for the differences in hydraulic potential ("head") that drive water through porous rock or sediments, analyze well-drilling cores to get a picture of underground rock and sediment layers, perform pump tests to determine how readily the water can flow through the strata, and probe water samples for their geochemistry and contamination.

His forte, though, is running the computer models that hydrogeologists use to integrate their knowledge of a groundwater system and plan its future. Zeiler's modeling work has covered sites in California, Montana, Alaska, and Ghana. His biggest project is an aquifer east of Los Angeles where groundwater contaminated with fuel and industrial solvents is being pumped out and treated for use as drinking water. Zeiler says he leaves most of the data gathering to other scientists. "I enjoy being outside—working on a drill rig, getting my hands dirty, all that stuff," he says. "But modeling is where I've ended up."

Taking data. Kurt Zeiler logs core samples.

Of the 15 hydrogeologists in the Denver office where Zeiler works, only three have Ph.D.s. That's typical of the field as a whole: AGI estimates that university programs graduate five times as many M.S. students as Ph.D.s. Its figures show that about 18,000 hydrologists and hydrogeologists now work in the environmental industry, a few thousand in the mining and petroleum industries, and about 850 in academia, the only sector for which a doctorate is required.

Despite high demand, salaries for hydrogeologists in government and in the private sector remain about 15% to 20% below those of other geoscientists. Low payoffs go hand in hand with high security, Haggerty says: "I know of people in their 50s who have been fabulously successful in the petroleum industry, but there are boom-and-bust cycles. In hydrogeology, the boom and bust is not there. It's much more level. But I don't know of many multimillionaire hydrogeologists, either."

If the field doesn't surge, it does at least ebb and flow. "Ten to 20 years ago, it was all cleanup—contaminant work," Allen-King says. The profession's cleanup phase waned in the late 1990s and early 2000s as changing political priorities, soaring cleanup costs, and some noteworthy environmental successes led to cutbacks in remediation.

Now, the focus has shifted to supply, the problem of finding and managing water resources while protecting their quality. Techniques such as artificial recharge (reinjecting water into the ground for storage) and carbon sequestration (keeping carbon dioxide emissions from combustion out of the atmosphere by forcing them underground) raise new water-related environmental issues that scientists are just learning to tackle.

In the realm of research, academic hydrogeologists are broadening their time horizons to help forecast and mitigate the effects of climate change, and they're stretching the traditional boundaries of their field to explore questions such as how groundwater interacts with the surface water of lakes and rivers. Collaborations with scientists from other disciplines are proliferating. "More and more, hydrogeologists are no longer working alone," says John Wilson, a hydrology professor at the New Mexico Institute of Mining and Technology in Socorro. "The subsurface is part of it but no longer the whole thing. Other parts of the cycle—biology, chemistry, ecology, and climate—that is where research is going, I think."

CREDITS (TOP TO BOTTOM): TRAYLE KULSHAN; KURT ZEILER

Shades of green

The broadening of academic research has filtered down to the training of M.S. students. Some graduate programs, such as New Mexico Tech's and the double-M.S. program at the University of Wisconsin, now require interdisciplinary courses in topics such as surface water, ecology, and economics. Some hydrologic consulting companies, however, complain that versatility isn't what they need. "We're having difficulty finding traditional hydrogeologists," says Daniel Stephens, founder and head of a 110-employee environmental consulting company with offices in New Mexico, Texas, and California. "The people we're seeing are fewer in number, and their qualifications are thinner." Instead of giving students a smorgasbord of skills, Stephens says, universities should equip them to start work on real projects.

But Wilson, whose department at New Mexico Tech embraced the multidisciplinary approach a decade ago, says his students are well-prepared to learn anything they need to know. "At some point, the employer is responsible for training students in the details of the jobs," he says. Oil companies, he notes, are happy to recruit promising hydrogeologists and train them in petroleum exploration.

If money were the only lure for earth scientists, fossil fuels might be the only fluids in town. But hydrogeologists say a strong undercurrent of environmental idealism pervades the field as well. Just as many senior scientists drew their inspiration from the first Earth Day, some young water experts are pursuing activist agendas of their own—and carving out new career paths to do it.

Trayle Kulshan is one of them. After getting her M.S. in hydrogeology from Stanford University in Palo Alto, California, in 2002, she spent 2 years in the Peace Corps in Guinea. Now, as water, sanitation, and hygiene coordinator for the humanitarian nongovernmental organization Action Contre la Faim (Action Against Hunger), she plans and oversees projects to build wells, latrines, and water networks in developing countries as far-flung as Afghanistan and the Democratic Republic of the Congo. She currently is working in Kenya. Although she is now as much a public-health worker as an earth scientist, Kulshan says her graduate training gives her a quick grasp of hydrologic conditions as well as skill in collecting, analyzing, and interpreting other kinds of data and communicating the results. "In grad school, we are all [teaching assistants] and develop skills as teachers. Every day I am teaching," she wrote by e-mail from Nairobi. "And I have to say I am still a student learning every day as well."

–ROBERT COONTZ

continued from page 857

exploration companies such as Schlumberger and companies that integrate the entire process from exploration to production to distribution, such as BP and ExxonMobil.

Recruitment and preparation

Sally Serenyi didn't set out to join the oil industry. But not long after she graduated with her bachelor's degree in physics from the University of Exeter in the U.K., she walked past a glossy recruiting display at a career fair with pictures of exotic landscapes. The display was for Schlumberger; now, just under 2 years later, she works for the company as a field engineer. "I joined for a bit of adventure," she says, and the "opportunity to go all over the world." Serenyi works with a crew of a half-dozen engineers and equipment operators near exploration sites in Austria and Hungary, collecting and preparing data for client companies.

Recruiters generally encourage students interested in the oil industry to obtain master's degrees, and industry-sponsored scholarships provide some incentive. But the current demand for new talent means that companies are "happy to have people with a strong bachelor's," says Charles Groat, former USGS director and now a professor of energy and mineral resources at UT Austin. Groat tells students that oil companies still prefer graduates who have fundamental training in geology and geophysics and who are comfortable with quantitative analysis, perhaps through classes in economics, statistics, or computer science. There are also jobs available to people in related fields including physics and engineering, as Serenyi found.

Companies use internship programs as recruiting tools and as extended interviews, says Eric Lyons, a geophysicist at BP. Lyons did three internships with Marathon Oil Corp. en route to earning a geophysics master's degree. Such programs give companies "a chance to look at the students and vice versa," Groat says. But internships are "a long way from being required," he adds.

The daily grind

New recruits to large companies typically spend a significant part of the first couple of years completing in-company training and gaining experience with different parts of the company. Lyons, who has worked on North American oil fields, now works on fields in the Gulf of Mexico and will be assigned elsewhere next year. For DeMott, who earned her master's degree in geology in 2007, training at ExxonMobil will involve three 8-month placements.

As companies bring in green staff, they are also trying to bring experienced staff back from retirement, or retain baby boomers as part-time consultants, to train the young recruits. Even so, the workforce is expected to be bottom-heavy—which could work to the advantage of younger employees. "In a few years, there will be the option to go up the ladder faster," Diaz-Tushman says.

In large companies, young geoscientists often have a choice of pursuing a managerial track or a parallel technical career ladder, with comparable compensation and recognition. In smaller companies that have just one or two geologists on staff, that may not be possible, and the work can be "more mundane, since they don't have the resources for more exotic stuff," says Groat. Still, smaller companies can have a different culture that may suit some geoscientists, he adds.

Lyons warns students who "love being outside" that at most oil industry jobs, "you're gonna be sitting inside all the time." Diaz-Tushman says that even though some of her work is in the field, ultimately, "I have an office job." It is possible to find fieldwork-focused jobs, such as Serenyi's, but fieldwork can mean giving up sleep when a well needs attention or a client suddenly needs data. "[The lifestyle] wouldn't fit someone who wasn't particularly energetic," Serenyi says.

Challenges

Plenty of technical challenges await this new generation of geoscientists. "The hottest area right now is unconventional reservoirs," says Groat. These include tight reservoirs of traditional hydrocarbons that have no natural fractures for engineers to exploit as conduits. Oil companies are also exploring other hydrocarbon sources such as heavy oils, coal-bed methane, and oil shale.

"There's a lot of stuff out there," DeMott emphasizes, "but what's there is not easy and not cheap to get out; … that's the problem that everyone is facing."

It's hard to forget the price crashes, layoffs, and hiring freezes that swept the industry during the 1980s and 1990s. But analysts predict not just stability but strong growth in jobs in the field in the near future.

"Maybe the hiring rate will slow in 5 or 10 years," DeMott says, "but I'm not concerned with losing my job. There's still not that many people to hire."

–LUCAS LAURSEN

Lucas Laursen is a freelance science writer in Cambridge, U.K.

POSITIONS OPEN

OXYGEN SENSING and REDOX SIGNALING SCIENTIST

The Department of Medicine/Section of Cardiology is seeking qualified applicants for a full-time **RESEARCH ASSOCIATE (ASSISTANT/ASSOCIATE/PROFESSOR)** position to work with **Dr. Stephen Archer** on studies of basic mechanism of oxygen sensing and experimental therapies for pulmonary hypertension. The successful applicant will join the heart and vascular research group, led by Dr. Archer.

The primary activity of the Research Associate (Assistant/Associate/Professor) is academic research, along with the development of a redox signaling and reactive oxygen species (ROS) research core in association with a faculty member or team. Qualified applicants are required to possess a Doctorate degree in a relevant discipline (e.g. free radical biology and/or biochemistry). Applicants should possess excellent knowledge of free radical biology and redox signaling regulation and have a demonstrated track record of expertise performing in vivo and in vitro experiments with pharmacological or genetic manipulation of antioxidants and detection of ROS. Experience in small animal work (including small animal surgeries), usage of electron paramagnetic resonance spectroscopy (EPR), fluorescent/luminescent measurement equipment (including confocal microscopy), and molecular biology techniques preferred. Experience and the ability to manage various levels of personnel within a laboratory environment, assist with monitoring laboratory budget expenditures, and overall laboratory management also required. Three to five years of postdoctoral training is required. A demonstrated track record of publication and the potential to apply for peer-reviewed funding is preferred. Responsibilities include data management, analysis, manuscript preparation, and submission of protocols for approval by the Institutional Animal Care and Use Committee. Preference will be given to applicants who have experience at the University of Chicago. Compensation and level of appointment are dependent on qualifications. The University provides a generous package of fringe benefits. Interested applicants should submit cover letter, curriculum vitae, and three letters of reference via e-mail to **Dr. Stephen Archer, e-mail: sarcher@medicine.bsd.uchicago.edu.** *The University of Chicago is an Affirmative Action/Equal Opportunity Employer.*

The Department of Basic Sciences in the College of Veterinary Medicine (CVM) at Mississippi State University (MSU) is seeking applicants for a tenure-track **ASSISTANT PROFESSOR** in systems biology. Research will be a major component of this position, and evidence of or potential for extramural funding is necessary. Teaching in the graduate and professional curricula (depending on area of expertise) is also expected. Involvement in applied systems biology in animals and microbes of agricultural/veterinary interest is also needed to facilitate collaboration with other faculty in the CVM and MSU. The successful applicant will become a member of the Institute for Digital Biology (IDB), a vibrant university-wide community of researchers. The IDB fosters collaboration and provides access to leading edge computer facilities and expertise. Minimum qualifications for this position include a Ph.D. and/or postgraduate training in systems biology and/or a D.V.M. or other terminal degree with evidence of published research expertise in applied systems biology. The ability to work both independently and as a team member will be useful in this environment. Salary will be commensurate with experience and will be competitive. Evaluation of applications will begin July 30, 2008, and continue until the position is filled. Please send curriculum vitae, names and contact information for three persons who can provide letters of recommendation, and a brief (one page or less) statement of research plans and goals as well as teaching philosophy to: **Dr. Stephen B. Pruett, Department of Basic Sciences, College of Veterinary Medicine, P.O. Box 6100, Mississippi State, MS 39762.** *MSU is an Affirmative Action/Equal Opportunity Employer.*

POSITIONS OPEN

DIRECTOR
Glennan Center for Geriatrics and Gerontology
Eastern Virginia Medical School
Norfolk, Virginia

Eastern Virginia Medical School (EVMS) seeks applications for the position of Director of the Glennan Center for Geriatrics and Gerontology with a faculty appointment in the Department of Internal Medicine at the level of **ASSOCIATE PROFESSOR or PROFESSOR.** Candidates should have an M.D. or Ph.D. degree, must have demonstrated excellence in research, and possess exceptional leadership qualities. The Director will have the opportunity to build a prominent center, emphasizing excellence in research, teaching, and clinical care related to aspects of aging. EVMS is undergoing a significant expansion in the areas of basic and translational research. There are significant resources available, including excellent laboratory space and endowed professorships for recruitment of the Center Director and additional faculty.

The Glennan Center for Geriatrics and Gerontology has gained national and international recognition for excellence in immunology, driving, and cognition in the context of aging research. The Center is also a leader in clinical care, providing innovative services to meet the special health care needs of older adults across a full range of practice settings from independent living to assisted living, long-term care, palliative care, and hospital care. The Center currently consists of eleven faculty members, and offers a comprehensive program for clinicians and scientists that provides training in geriatrics and gerontology for medical students, residents, fellows, other health care professionals, and junior faculty members.

Eastern Virginia Medical School is located in coastal southeastern Virginia in the nation's 27th largest metropolitan statistical area. The region offers premier waterfront communities, large beaches, excellent golf, tennis, sailing, and other recreational opportunities, and top ranked schools.

Please send a letter of interest including current curriculum vitae to the **Executive Search Committee** at **e-mail: execcomm@evms.edu.** (*Affirmative Action-Equal Opportunity Employer.*)

FACULTY POSITION in PHYSICAL CHEMISTRY

Physical chemistry faculty position, University of Oregon. The Department of Chemistry invites applications for a tenure-track position in experimental physical chemistry (broadly defined) beginning fall 2009 at the anticipated rank of **ASSISTANT PROFESSOR.** A Ph.D. is required and postdoctoral experience is preferred. Individuals studying problems of basic importance in experimental physical chemistry are especially encouraged to apply. The potential for establishing a vigorous independent research program in physical chemistry and active participation and excellence in teaching at the undergraduate and graduate levels will be the most important criteria for selection. In addition to membership in the Department of Chemistry, the successful candidate will have the opportunity to pursue research in a variety of interdisciplinary institutes. We seek candidates who will function with high standards of professionalism with students, faculty, and staff from diverse backgrounds. Applications will be accepted until the position is filled. To assure full consideration, applicants should submit curriculum vitae and a statement of research plans and objectives by October 15, 2008, and arrange for three letters of recommendation to be sent to: **Physical Chemistry Search Committee, 1253 Department of Chemistry, University of Oregon, Eugene, OR 974030-1253.** *The University of Oregon is an Equal Opportunity/Affirmative Action/ADA Institution committed to cultural diversity.*

POSITIONS OPEN

ENVIRONMENTAL CHEMISTRY. Haverford College seeks to fill an open-rank, tenure-track position in environmental chemistry, broadly defined. Applicants from all areas of chemistry are encouraged to apply, but expertise should focus on topics related to the environment. The successful candidate will teach chemistry courses at all undergraduate levels and play a formative role in a new interdisciplinary program in environmental studies. Candidates must have a strong commitment to teaching and the establishment of a vigorous research program involving undergraduate students. A Ph.D. and postdoctoral experience are required. Applicants should submit curriculum vitae, with statements of research plans and teaching philosophy electronically to **e-mail: hc-chemistrysearch@haverford.edu.** Undergraduate and graduate transcripts, and three letters of recommendation should be mailed to: **Merleen Macdonald, Search Administrative Assistant; Haverford College, 370 Lancaster Avenue, Haverford, PA 19041.** Questions should be directed to: **Dr. R. Scarrow, Professor of Chemistry (e-mail: rscarrow@haverford.edu).** For full consideration, applications must be completed by October 10, 2008. More information is available from **website: http://www.haverford.edu/chemistry.**

Haverford College **(website: http://www.haverford.edu)** is located 12 miles northwest of Philadelphia in the Delaware River watershed, an area of intense industrial, urban, and agricultural activity which also includes many parks and protected natural areas. *Haverford is an Equal Opportunity/Affirmative Action Employer, committed to excellence through diversity, and strongly encourages applications and nominations of persons of color, women, and members of other underrepresented groups.*

FACULTY POSITION
Chemical Biology and Bioorganic Chemistry
University of California, Irvine

The Department of Chemistry of the University of California, Irvine invites applications for a tenure-track position at the **ASSISTANT PROFESSOR** level in chemical biology and bioorganic chemistry. We seek a Ph.D.-level scientist who will establish a vigorous research program in chemical biology broadly defined. The candidates should also be committed to teach chemistry at the undergraduate and graduate levels. Applicants should send their curriculum vitae, a list of publications, and a description of their proposed research program, to the: **Chemical Biology Search Committee, Department of Chemistry, University of California, Irvine, CA 92697-2025.** Applications may also be submitted electronically via the web at **website: https://recruit.ap.uci.edu/.** Web applications should include a cover letter and the information requested above. Applicants should also arrange to have three letters of recommendation submitted on their behalf. To insure full consideration, applications and supporting materials should be received by October 31, 2008. *The University of California is an Equal Opportunity/Affirmative Action Employer committed to excellence through diversity, and has an ADVANCE program dedicated to gender and ethnic equity.*

ASSISTANT PROFESSOR
The University of Chicago
Department of Chemistry

The Department of Chemistry of the University of Chicago invites applications from outstanding individuals for the position of Assistant Professor of Chemistry. This search is in the areas broadly defined as inorganic, organic, and physical chemistry. Applications must include curriculum vitae, a list of publications, and a succinct outline of research plans, and be supported by three letters of recommendation. Materials accepted by mail only addressed, as appropriate, to: **Inorganic Search Committee, Organic Search Committee, or Physical Search Committee, Department of Chemistry, Office of the Chairman (GHJ 222), The University of Chicago, 5735 S. Ellis Avenue, Chicago, IL 60637.** Review of completed applications will begin October 1, 2008; to ensure full consideration, all material should be submitted by that date.

University of Chicago is an Affirmative Action/Equal Opportunity Employer.

POSITIONS OPEN

Society for Conservation Biology (SCB) seeks new **CONSERVATION BIOLOGY EDITORS**.

SCB is initiating a change in the editorial structure of its internationally acclaimed journal *Conservation Biology*, and seeks a diverse and international team of five Editors, one of them **EDITOR-IN-CHIEF**, to lead the journal. We invite applications from individuals with experience in editing scientific publications and expertise in conservation science, understood as a discipline that spans the social and natural sciences, as well as across biological realms and geographical regions.

Compensation will be in the form of honoraria, and the appointments will be for up to four years, renewable. The team of Editors will be supported by experienced, professional staff already employed by SCB.

Please send expressions of interest and resume, by 15 September 2008, to: **Chair, Conservation Biology Editors Search Committee, e-mail: editorsearch@conbio.org.** Additional details of duties and terms of employment may be found at **website: http://www.conbio.org/EditorSearch.**

The ENDOWMENT for SCHOLARS in BIOMEDICAL RESEARCH

The University of Texas Southwestern Medical Center UT Southwestern is pleased to announce the continuation of the Endowed Program for Scholars in Biomedical Research. The Program, which is fully funded from private endowment, will provide at least $1,000,000 over four years to support the research activities of each new **ASSISTANT PROFESSOR** (tenure track) appointed to the Program; five will be appointed annually. Academic appointments and research space will be provided by individual medical school departments or research centers. Positions in both basic science and clinical departments are available. The goal of the program is to assure a successful beginning of the research careers of an evergrowing cadre of outstanding young investigators at UT Southwestern.

For detailed information about currently available faculty positions, please access our **website: http://www8.utsouthwestern.edu/utsw/home/scholars/index.html.**

SOUTHWESTERN
The University of Texas Southwestern Medical Center at Dallas

UT Southwestern is an Equal Opportunity Institution.

POSTDOCTORAL POSITIONS. Two positions are available to investigate unchartered biochemistry and physiology of anaerobic microbes that also addresses a broader understanding of diverse prokaryotes from both the *Archaea* and *Bacteria* domains. The successful candidates will have the opportunity to apply a diversity of experimental approaches, depending on experience and objectives for advanced training, which include crystallography, spectroscopy, enzymology, genetics, and genomic expression profiling. The laboratory is generously supported by National Science Foundation, Department of Energy, NASA, Air Force Office of Scientific Research, and Pennsylvania State University. For more information visit **website: http://www.bmb.psu.edu/faculty/ferry/lab/index.html.** Send resume and names of three references to: **James G. Ferry, Job #A-28403, 205 S. Frear Laboratory, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802. E-mail: jgf3@psu.edu.** Materials accepted until positions are filled. *Penn State is committed to Affirmative Action, Equal Opportunity, and the diversity of its work force.*

POSITIONS OPEN

The Department of Biology at the Citadel, the Military College of South Carolina, invites applications for a tenure-track **GENETICIST** at the **ASSISTANT PROFESSOR** level beginning January 2009 (an August 2009 starting date is also possible). Candidates should possess a Ph.D. in genetics or a related area. Primary teaching responsibilities will include sophomore and graduate level genetics courses for biology majors, upper-division and graduate courses in molecular genetics and the candidate's area of expertise, and participation in the Department's introductory biology courses. The Department offers a B.S. biology major and a core curriculum science sequence for the Corps of Cadets, and graduate programs leading to the M.A. and M.A.T. degrees. Candidates will be expected to develop an active research program. The Charleston area offers ample opportunities for collaborative research with state and federal agencies, including the Medical University of South Carolina. Citadel faculty may also serve as adjunct faculty for the College of Charleston's Graduate Programs in Marine Biology and Environmental Studies. Interested candidates should send a letter of application; curriculum vitae; statements of teaching and research interests; and names, telephone numbers, and e-mail addresses of three references to: **Dr. Paul Rosenblum, Professor and Head, Department of Biology, The Citadel, 171 Moultrie Street, Charleston, SC 29409. Fax: 843-953-5228.** You may also submit application and materials online at **website: http://www.citadel.edu/hr.** Please reference job #F08-11. Additional information is available at the Department of Biology **website: http://www.citadel.edu/citadel/otherserv/biol/.**

The Citadel is an Affirmative Action/Equal Opportunity Employer actively committed to ensuring diversity in all campus employment. (006110)

Two **POSTDOCTORAL POSITIONS in MICROBIAL IMMUNOLOGY and PATHOGENESIS.** University of Wisconsin, Madison. One position addresses host-microbe interactions and seeks to identify new innate receptors for fungi. The work explores how mammalian receptors for fungi trigger immunoregulatory disturbances and inflammatory diseases such as asthma, a global health problem. Cell and molecular approaches will address known receptors and discover new ones. Background and experience in cell and molecular biology desirable. A second position addresses structure: function relatioships of a virulence factor in *Blastomyces* dermatitidids, how it lodges on the fungal surface, interacts with and modulates the host. Background and experience in molecular biology and biochemistry or biophysics desirable. Madison is one of the most livable U.S. cities. The laboratory is located in a new, state-of-the-art building, with many opportunities for scientific collaboration. The positions are funded by NIH or the Sandler Foundation. Resume to: **Bruce Klein, University of Wisconsin-Madison, Room 4303 Microbial Sciences Building, 1550 Linden Drive, Madison, WI 53706-1521; e-mail: bsklein@wisc.edu. Telephone: 608-263-9217.**

POSTDOCTORAL FELLOW

Postdoctoral positions are available immediately in the laboratories of **Drs. Qiangrong Liang and A. Martin Gerdes** at the Cardiovascular Research Center at Sanford Research/University of South Dakota. Fellows will study molecular and cellular mechanisms associated with cardiac hypertrophy and/or angiogenesis using in vitro and in vivo mouse models. Candidates should have a Ph.D., M.D./Ph.D., or D.V.M./Ph.D. degree(s). A strong background in cardiovascular biology is preferred but not required. For the complete job description and online application see **website: http://www.sanfordhealth.org.** Please refer to requisition number #16326 and/or #16264.

Sanford
1305 W. 18th Street
Sioux Falls, SD 571170-5039
Telephone: 605-333-7000
Website: http://www.sanfordhealth.org

Equal Opportunity Employer.

POSITIONS OPEN

POSTDOCTORAL POSITIONS

Durham, North Carolina

Postdoctoral positions are available immediately to investigate anti-tumor immune response induced by high-intensity focused ultrasound (HIFU) and to develop novel technologies to combine synergistically HIFU with immunotherapy for prostate, breast, and skin cancer therapy **(website: http://www.dukenews.duke.edu/2007/08/hifu.html).** The successful candidate will join a multidisciplinary team of investigators from the Engineering School and Duke Comprehensive Cancer Center. Highly motivated candidates with cancer immunology and surgery expertise are desirable. Please send curriculum vitae and three reference letters to: **Pei Zhong, Ph.D., e-mail: pzhong@duke.edu.**

POSTDOCTORAL RESEARCH ASSOCIATE POSITIONS in NANOTECHNOLOGY at Oakland University, Rochester, Michigan. The NanoTech R&D Institute (Nano Institute) at Oakland University **(website: http://www.oakland.edu)** has immediate openings for Postdoctoral Associates with expertise in the general area of nanotechnology. The Nano Institute has active research collaboration with a local nanotechnology incubator and other industrial partners. Applicants must have an earned Ph.D. degree in science, engineering, or other related fields. Preference will be given to applicants with photovoltaic and/or other renewable energy experiences. All positions are initially for one year and renewable for additional years. Interested candidates should e-mail complete curriculum vitae, research statements, and contact information for three references to **e-mail: yih@oakland.edu.**

POSTDOCTORAL POSITIONS are available in Institute for Cancer Genetics at Columbia University, to study protein modifications in cancer and aging (e.g., **Zhao et al.,** *Nature* **451:**587, 2008; **Tang et al.,** *Cell* **133:**612, 2008; and **Brooks and Gu,** *Mol. Cell* **21:**307, 2006). Prior background and fresh graduates in related areas would be helpful but not essential. Candidates should have Ph.D. or M.D. and must be highly motivated.

Send curriculum vitae and names of three references to **Ms. Kimberly Labor/Dr. Wei Gu** at **e-mail: kjl2006@columbia.edu** or **Insititute for Cancer Genetics, Room #609, Columbia University, 1130 Street Nicholas Avenue, New York, NY 10032 USA.**

Columbia University is an Affirmative Action/Equal Opportunity Employer.

LABORATORY POSITION AVAILABLE

The Vascular Biology Laboratory at the Providence VA Medical Center is seeking a full-time **RESEARCH TECHNICIAN or POSTDOCTORAL FELLOW** to begin work immediately on an NIH-funded grant that aims to determine the role of natriuretic peptides in modulating pulmonary endothelial barrier function. Complete job requirements are posted at **website: http://www.brown.edu/Research/Vascular_Research_Laboratory/Jobs.html.**

Interested individuals should contact: **Dr. James Klinger, Associate Professor, Brown Medical School, Division of Pulmonary/Critical Care Medicine, Rhode Island Hospital, 593 Eddy Street, Providence, RI 02903** or **e-mail: james_klinger@brown.edu.**

POSTDOCTORAL POSITION is immediately available in the area of cardiovascular engineering to develop multiscale computational models for (1) embryonic heart development, and (2) percutaneous heart valve systems. Applicants should hold a Ph.D. in mechanics or related disciplines, expert in numerical analysis, computational methods/software, and 3D-CAD design software. For more information please e-mail curriculum vitae to **Prof. Kheradvar (e-mail: arashkh@sc.edu)** at University of South Carolina *(Equal Opportunity Employer).*

POSITIONS OPEN

UNIVERSITY of HAWAII at HILO

ASSISTANT PROFESSOR of BIOLOGY. Position number 82909, University of Hawai'i at Hilo, College of Arts and Sciences, full-time, nine-month, tenure-track appointment to begin in 2009.

Duties: Teach undergraduate courses in introductory biology (including botany); teach and participate in the Tropical Conservation Biology and Environmental Science Graduate Program.

For a complete description and application requirements, please go to **website: http://www.uhh.hawaii.edu/uhh/hr/jobs.php.**

The University of Hawaii is an Equal Employment Opportunity/Affirmative Action Employer. Persons with Disabilities/Minorities/Veterans/Women.

MARKETPLACE